Locally Decodable Codes and Private Information Retrieval Schemes

by

Sergey Yekhanin

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

July 2007
[September 2007]

Author ....
Department of Electrical Engineering and Computer Science
July 2, 2007

Certified by ....
Madhu Sudan
Fujitsu Professor of EECS, MIT
Thesis Supervisor

Accepted by ....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Locally Decodable Codes and Private Information Retrieval Schemes

by

Sergey Yekhanin

Submitted to the Department of Electrical Engineering and Computer Science
on July 2, 2007, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis studies two closely related notions, namely Locally Decodable Codes (LDCs) and Private Information Retrieval Schemes (PIRs).

Locally decodable codes are error-correcting codes that allow extremely efficient, "sublinear-time" decoding procedures. More formally, a k-query locally decodable code encodes n-bit messages x in such a way that one can probabilistically recover any bit x_i of the message by querying only k bits of the (possibly corrupted) codeword, where k can be as small as 2. LDCs were initially introduced in complexity theory in the context of worst-case to average-case reductions and probabilistically checkable proofs. Later they have found applications in numerous other areas including information theory, cryptography and the theory of fault tolerant computation. The major goal of LDC related research is to establish the optimal trade-off between length N and query complexity k of such codes, for a given message length n.

Private information retrieval schemes are cryptographic protocols developed in order to protect the privacy of the user's query, when accessing a public database. In such schemes a database (modelled by an n-bit string x) is replicated between k non-communicating servers. The user holds an index i and is interested in obtaining the value of the bit x_i. To achieve this goal, the user queries each of the servers and gets replies from which the desired bit x_i can be computed. The query to each server is distributed independently of i and therefore each server gets no information about what the user is after. The main parameter of interest in a PIR scheme is its communication complexity, namely the number of bits exchanged by the user accessing an n-bit database and the servers.

In this thesis we provide a fresh algebraic look at the theory of locally decodable codes and private information retrieval schemes. We obtain new families of LDCs and PIRs that have much better parameters than those of previously known constructions. We also prove limitations of two server PIRs in a restricted setting that covers all currently known schemes. Below is a more detailed summary of our contributions.

- Our main result is a novel (point removal) approach to constructing locally decodable codes that yields vast improvements upon the earlier work. Specifically, given any Mersenne prime $p = 2^t - 1$, we design three query LDCs of length $N = \exp\left(n^{1/t}\right)$, for every n. Based on the largest known Mersenne prime, this translates to a length of less than $\exp\left(n^{10^{-7}}\right)$, compared to $\exp\left(n^{1/2}\right)$ in the previous constructions. It has often been conjectured that there are infinitely many Mersenne primes. Under this conjecture, our constructions yield three query locally decodable codes of length $N = \exp\left(n^{O\left(\frac{1}{\log\log n}\right)}\right)$ for infinitely many n.

- We address a natural question regarding the limitations of the point-removal approach. We argue that further progress in the unconditional bounds via this method (under a fairly broad definition of the method) is tied to progress on an old number theory question regarding the size of the largest prime factors of Mersenne numbers.

- Our improvements in the parameters of locally decodable codes yield analogous improvements for private information retrieval schemes. We give 3-server PIR schemes with communication complexity of $O\left(n^{10^{-7}}\right)$ to access an n-bit database, compared to the previous best scheme with complexity $O(n^{1/5.25})$. Assuming again that there are infinitely many Mersenne primes, we get 3-server PIR schemes of communication complexity $n^{O(1/\log\log n)}$ for infinitely many n.

- Our constructions yield tremendous improvements for private information retrieval schemes involving three or more servers, and provide no insights on the two server case. This raises a natural question regarding whether the two server case is truly intrinsically different. We argue that this may well be the case. We introduce a novel combinatorial approach to PIR and establish the optimality of the currently best known two server schemes a restricted although fairly broad model.

Thesis Supervisor: Madhu Sudan
Title: Fujitsu Professor of EECS, MIT

Acknowledgments

I have greatly benefited from Vyacheslav V. Rykov and Arkadii G. D'yachkov, who inspired my interest in mathematics and computer science while I was in high-school. I am indebted to them for getting me involved in coding theory research during the first years of my undergraduate studies in Moscow State University and being my well-wishers ever since.

I am greatly indebted to Madhu Sudan, my advisor for the years of careful nurturing that he provided to me. Madhu's deep intuition and creative approach to technical issues have influenced me a lot. Madhu's highly insightful comments regarding the problems I worked on have been a big factor in aiding my thinking.

While at MIT, I was extremely lucky to meet Swastik Kopparty, my good friend, who provided me with an immense amount of technical expertise in many areas of mathematics.

I am indebted to Kiran Kedlaya, Alexander Razborov, and David Woodruff for their collaborations that led to several of the results discussed in this thesis. I also wish to thank the several other people with whom I have had useful discussions on the topics related to this thesis. These include Noga Alon, Ilya Dumer, Oded Goldreich, Yuval Ishai, Swastik Kopparty, Igor Pak, Alex Samorodnitsky, Madhu Sudan, Sasha Postnikov, Alex Vardy, and Avi Wigderson.

I would like to thank Cynthia Dwork and all the members of the theory group in Microsoft Research Silicon Valley Lab – Steve Chien, Frank McSherry, Ilya Mironov, and Kunal Talwar - who hosted me during a fun and productive summer visit in 2006.

The academic environment at MIT has on one hand been very stimulating and on the other it has provided me with a wonderful set of friends Alex Andoni, Elena Grigorescu, MohammadTaghi Hajiaghayi, Nick Harvey, Jon Kelner, Swastik Kopparty, Sayan Mitra, Eddie Nikolova, Mihai Patrascu, Tasos Sidiropoulos, David Woodruff, Hanson Zhou - thank you!

To my sister Yulia

Contents

Chapter 1

Introduction

This thesis studies two closely related notions, namely Locally Decodable Codes (LDCs) and Private Information Retrieval Schemes (PIRs). Locally decodable codes are error-correcting codes that allow extremely efficient, sublinear-time decoding procedures. Private information retrieval schemes are cryptographic protocols developed in order to protect the privacy of the user accessing a public database. We address a long-open question regarding the parameters of optimal LDCs and PIRs.

1.1 Locally decodable codes

LDCs are a special kind of error-correcting codes. Error-correcting codes are used to ensure reliable transmission of information over noisy channels as well as to ensure reliable storage information on a medium that may be partially corrupted over time (or whose reading device is subject to errors). In both of these applications the message is typically partitioned into small blocks and then each block is encoded separately. Such encoding strategy allows efficient random-access retrieval of the information, since one needs to decode only the portion of data one is interested in. Unfortunately, this strategy yields very poor noise resilience, since in case even a single block (out of possibly tens of thousands) is completely corrupted some information is lost. In view of this limitation it would seem preferable to encode the whole message into a single codeword of an error-correcting code. Such solution clearly improves the robustness to noise, but is also hardly satisfactory, since one now needs to look at the whole codeword in order to recover any particular bit of the message (at least in the case when classical error-correcting codes are used). Such decoding complexity is prohibitive for modern massive data-sets.

Locally decodable codes are error-correcting codes that avoid the problem mentioned above by having extremely efficient *sublinear-time* decoding algorithms. More formally, a k-query locally decodable code C encodes n-bit messages x in such a way that one can probabilistically recover any bit x_i of the message by querying only k bits of the (possibly corrupted) codeword $C(x)$, where k can be as small as 2.

The classical Hadamard code [MS77] provides the simplest nontrivial example of LDCs. The Hadamard code is a 2-query locally decodable code that encodes n-bit messages x to $\exp(n)$[1]-bit codewords $C(x)$. The key property of such an encoding is that given a codeword $C(x)$ corrupted in up to δ fraction of (adversarial chosen) locations and an integer i between 1 and n one can toss some random coins, query two locations of the corrupted codeword and output the value that agrees with x_i with probability $1 - 2\delta$, where the probability is over the random coin tosses only. The (local) decoding procedure described above allows a fairly reliable extremely efficient recovery of bits of x, as long as the fraction of corrupted locations is not too large.

The main parameters of interest in LDCs are the (codeword) length and the query complexity. The length of the code measures the amount of redundancy that is introduced into the message by the encoder. The query complexity counts the number of bits that need to be read from the (corrupted) codeword in order to recover a single bit of the message. Ideally, one would like to have both of these parameters as small as possible. One however can not minimize the length and the query complexity simultaneously. There is a trade-off. On one end of the spectrum we have classical error correcting codes [MS77, vL82] that have both query complexity and codeword length proportional to the message length. On the other end we have the Hadamard code that has query complexity 2 and codeword length exponential in the

[1] Throughout the thesis we use the standard notation $\exp(x) = 2^{O(x)}$.

message length. Establishing the optimal trade-off between the length and the query complexity is major goal of research in the area of locally decodable codes.

Interestingly, the natural application of locally decodable codes to data transmission and storage described above is neither historically earliest nor the most important. LDCs have a host of applications in other areas including cryptography [CGKS98, IK04], complexity theory [Tre04] and the theory of fault tolerant computation [Rom06]. In most of those applications one is interested in codes with extremely low (ideally, constant) query complexity. For this reason a lot of work on locally decodable codes had concentrated on this regime. We follow the same tradition. In this thesis we study the trade-off between the length and the query complexity of LDCs with emphasis on constant-query codes.

We conclude the section by presenting a concrete family of k-query LDCs that encode n long messages to $\exp(n^{1/(k-1)})$ long codewords for every n. This family is a classical example of locally decodable codes that can be traced back to [BFLS91, Sud92, PS94]. Our presentation assumes that the reader is familiar with some basic algebraic concepts such as finite fields and polynomials [LN83, dW03].

A locally decodable code based on polynomial interpolation. Fix $k \geq 2$ to be the desired query complexity. Pick a prime power $r \geq k+1$. Also pick an integer $m \geq k-1$, and set $n = \binom{m}{k-1}$. We now show how to construct a k-query LDC encoding n-long messages over the finite field $\mathbb{F}_r$ to r^m-long codewords (over the same field). Note that for fixed k and growing m this yields codes of length $\exp(n^{1/(k-1)})$.

We start with some notation. For a positive integer s let $[s]$ denote the set $\{1, \ldots, s\}$. Let $\gamma : [n] \rightarrow \{0,1\}^m$ be a bijection between the set $[n]$ and the set of m-long $\{0,1\}$-vectors of Hamming weight $k-1$. (The Hamming weight of a vector is simply the number of its nonzero coordinates.) Finally, for $i \in [n]$ and $j \in [k-1]$ let $\gamma(i)_j$ denote the j-th nonzero coordinate of $\gamma(i)$.

Now we define the encoding procedure. Given a message $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}_r^n$ consider a multivariate polynomial D in the ring $\mathbb{F}_r[x_1, \ldots, x_m]$,

$$D(x_1, \ldots, x_m) = \sum_{i=1}^{n} \alpha_i x_{\gamma(i)_1} \ldots x_{\gamma(i)_{k-1}}.$$

The key properties of the polynomial D are the following: on one hand D encodes the message (for every $i \in [n]$ we have $D(\gamma(i)) = \alpha_i$); on the other hand D has low degree ($\deg D = k - 1$). We define the encoding of our message to be the evaluation of the polynomial D on the whole space $\mathbb{F}_r^m$. Clearly, such an encoding has length r^m.

It remains to show that the code defined above is locally decodable. Assume we are given an evaluation of the polynomial D over $\mathbb{F}_r^m$ that is adversatively corrupted in a δ fraction of locations. We are also given some index $i \in [n]$ and want to recover $D(W)$, where $W = \gamma(i)$ by probing the (corrupted) evaluation of D in at most k points. Note that it does not make much sense to probe the value of D at the point W itself, since the value there may be corrupted. Instead, we pick a uniformly random vector $V \in \mathbb{F}_r^m$ and:

1. Probe the evaluation of D at points $W + \lambda V$ for k distinct nonzero $\lambda \in \mathbb{F}_r$;

2. Interpolate a degree $k - 1$ univariate polynomial $D(W + \lambda V) \in \mathbb{F}_r[\lambda]$ using the values obtained above to get $D(W)$.

Observe that each individual point where we probe D in the decoding process is uniformly random. Therefore with probability at least $1 - k\delta$ we never probe D at a corrupted point, and decode correctly.

1.2 Private information retrieval schemes

Consider a user that makes a query to a database. In many cases the user may want to keep private the identity of the database record that he is interested in. A good example is an investor that queries the stock-market database for the value of a certain stock. Such an investor may wish not to disclose the identity of the stock he is curious about. Private information retrieval (PIR) schemes are cryptographic protocols that enable users to retrieve records from public databases, while keeping private the identity of the retrieved records.

For a reader who has not come across the notion of private information retrieval schemes before, it may seem quite puzzling how one can retrieve database records without revealing their identity to the server holding the database. Note however that there is a trivial solution. Namely, whenever the user wants a single record, the user can ask for the copy of the whole database. This solution involves a tremendous communication overhead, and is clearly practically unacceptable. Unfortunately, it can be shown [CGKS98] that if the user wants to keep its privacy fully-protected (in the information theoretic sense), then the trivial solution that we mentioned is essentially optimal.

Interestingly, the negative result above only applies to databases that are stored on a single server (rather than being replicated across several servers). In a seminal paper, Chor et al. [CGKS98] came up with PIR schemes that enable private retrieval of records from replicated databases, with a nontrivially small amount of communication. In such protocols the user makes queries to each server holding the database. The protocol ensures that every individual server (by observing only the query sent to him) gets no information about the identity of the item the user is interested in.

Before going further, let us make the notion of private information retrieval schemes more concrete. We model database as an n-bit string x that is replicated between k non-communicating servers. The user holds an index i (which is an integer between 1 and n) and is interested in obtaining the value of the bit x_i. To achieve this goal, the user tosses some random coins, queries each of the servers and gets replies from which the desired bit x_i can be computed. The query to each server is distributed independently of i and therefore each server gets no information about what the user is after. The main parameters of interest in a PIR scheme are the number k of servers involved, and the communication complexity, namely the number of bits exchanged by the user accessing an n-bit database and the servers. The major goal of PIR related research to design k-server PIR schemes with optimal (i.e., the smallest possible) amount of communication for every k. In this thesis we work towards this goal, and obtain both positive and negative results regarding the communication complexity of private information retrieval schemes.

We conclude the section by presenting a concrete k-server PIR scheme due to David Woodruff and the author [WY05]. The scheme involves $O(n^{1/(2k-1)})$ communication to access an n-bit database and is arguably the most intuitive among currently known nontrivial PIR schemes. Our presentation assumes that the reader is familiar with some basic algebraic concepts such as polynomials, finite fields, and derivatives [LN83, dW03].

A PIR scheme based on polynomial interpolation. Fix $k \geq 1$ to be the desired number of servers. Pick a prime power $r \geq k+1$. Also pick an integer $m \geq 2k-1$, and set $n = \binom{m}{2k-1}$. In what follows we obtain a k-server PIR scheme with $O(m)$ bits of communication to access an n sized database with entries from the finite field $\mathbb{F}_r$. Note that for fixed k and growing m this yields schemes with $O(n^{1/(2k-1)})$ communication.

Following the technique from the example in section 1.1, pick $\gamma : [n] \to \{0,1\}^m$ to be a bijection between the set $[n]$ and the set of m-long $\{0,1\}$-vectors of Hamming weight $2k-1$. For $i \in [n]$ and $j \in [2k-1]$ let $\gamma(i)_j$ denote the j-th nonzero coordinate of $\gamma(i)$. Given a database $(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}_r^n$ each server obtains the following polynomial D in the ring $\mathbb{F}_r[x_1, \ldots, x_m]$,

$$D(x_1, \ldots, x_m) = \sum_{i=1}^{n} \alpha_i x_{\gamma(i)_1} \ldots x_{\gamma(i)_{2k-1}}.$$

Observe that for every $i \in [n]$ we have $D(\gamma(i)) = \alpha_i$, and $\deg D = 2k-1$. The basic idea behind our PIR scheme is the idea of polynomial interpolation. The user holds a point $W = \gamma(i) \in \mathbb{F}_r^m$ and wants compute $D(W)$ without revealing the identity of W to the servers. To this end the user randomly selects an affine line $L \in \mathbb{F}_r^m$ containing the point W and discloses certain points on L to the servers. Each server computes and returns the value of D and the values of partial derivatives of D at the point that it is given. Finally, the user reconstructs the restriction of D to L. In particular the user obtains the desired value of $D(W)$. Below is a more formal description.

We use the standard mathematical notation $\left.\frac{\partial D}{\partial x_l}\right|_Q$ to denote the value of the partial derivative of D with respect to x_l at point Q. Let $\lambda_1, \ldots, \lambda_k \in \mathbb{F}_r$ be distinct and nonzero. Let $\mathcal{U}$ denote the user and $\mathcal{S}_1, \ldots, \mathcal{S}_k$ denote the servers.

$\mathcal{U}$	:	Picks $V \in \mathbb{F}_r^m$ uniformly at random.
$\mathcal{U} \to \mathcal{S}_h$	:	$W + \lambda_h V$
$\mathcal{U} \leftarrow \mathcal{S}_h$	:	$F(W + \lambda_h V), \left.\frac{\partial F}{\partial x_1}\right|_{W+\lambda_h V}, \ldots, \left.\frac{\partial F}{\partial x_m}\right|_{W+\lambda_h V}$

Note that in the protocol above the input of each $\mathcal{S}_h$ is a uniformly random point in $\mathbb{F}_r^m$. Therefore our protocol is private. It is also easy to verify both the queries that the user sends to servers and the servers' responses are of length $O(m) = O(n^{1/(2k-1)})$. (Every query is simply a point in $\mathbb{F}_r^m$. Every response is a list of m values of partial derivatives of D plus the value of D itself.) It remains to show how the user obtains $D(W)$ from the servers' responses.

Consider the line $L = \{W + \lambda V \mid \lambda \in \mathbb{F}_r\}$. Let $d(\lambda) = D(W + \lambda V)$ be the restriction of D to L. Clearly, $d(\lambda_h) = F(P + \lambda_h V)$. Thus the user knows the values $\{d(\lambda_h)\}$ for all $h \in [k]$. However, the values $\{d(\lambda_h)\}$ do not suffice to reconstruct the polynomial d, since the degree of d may be up to $2k - 1$. The main observation underlying our protocol is that knowing the values of partial derivatives $\left.\frac{\partial D}{\partial x_1}\right|_{W+\lambda_h V}, \ldots, \left.\frac{\partial D}{\partial x_m}\right|_{W+\lambda_h V}$, the user can reconstruct the value of $d'(\lambda_h)$. The proof is a straightforward application of the chain rule:

$$\left.\frac{\partial d}{\partial \lambda}\right|_{\lambda_h} = \left.\frac{\partial D(W + \lambda V)}{\partial \lambda}\right|_{\lambda_h} = \sum_{l=1}^{m} \left.\frac{\partial D}{\partial x_l}\right|_{W+\lambda_h V} V_l.$$

Thus the user can reconstruct $\{d(\lambda_h)\}$ and $\{d'(\lambda_h)\}$ for all $h \in [k]$. Combining this observation with the standard algebraic fact that a univariate polynomial of degree $2k - 1$ is uniquely defined by its values and derivatives at k points [LN83], we conclude that the user can reconstruct d and obtain $D(W) = d(0)$.

1.3 The history of LDCs and PIRs

Both locally decodable codes and private information retrieval schemes can be seen as the combinatorial analogs of notions that had been studied in complexity theory in the late 1980s and early 1990s. In particular, decoding procedures of locally decodable codes can be seen as combinatorial version of self-correctors [Lip90, BLR93], and private information retrieval schemes are analogous to instance-hiding schemes [AFK89, BF90, BFKR90]. Private information retrieval schemes were introduced by Chor et al. [CGKS98] in 1995. Locally decodable codes were explicitly discussed in the PCP literature in early 1990s, most notably in [BFLS91, Sud92, PS94]. However the first formal definition of LDCs was given only in 2000 by Katz and Trevisan [KT00], who also recognized an intimate relationship between LDCs and PIRs. (Namely, that LDCs yield PIRs with related parameters and vice versa.) Since then the study of LDCs and PIRs has grown into a fairly broad field, with many connections to other areas of theoretical computer science. There are two (somewhat outdated) surveys of LDC/PIR literature available [Tre04, Gas04].

One can informally classify the known families of LDCs and PIRs into three generations based on the technical ideas that underline these constructions. The latest (third) generation is the main contribution of this thesis. In the following sections we review every generation of codes and schemes and then proceed to lower bounds.

1.3.1 The first generation of LDCs and PIRs: interpolation

The first generation of LDCs and PIRs captures codes and schemes that are based on the idea of (low-degree) multivariate polynomial interpolation. Examples of such codes and schemes were given in sections 1.1 and 1.2.

All locally decodable codes of the first generation [KT00, BIK05] are variants of the classical Reed-Muller codes [MS77, vL82] that were introduced in coding theory in 1960s. For constant query complexity $k \geq 2$, locally decodable codes of the first generation have length $\exp(n^{1/(k-1)})$.

There is a number of different PIR schemes of the first generation. The earliest such schemes are due to Chor et al. [CGKS98]. Those schemes have communication complexity $O(n^{1/3})$ for the case of two servers, and communication complexity $O(n^{1/k})$ for the general k-server case. Later, different PIR schemes of the first generation were proposed by Ambainis [Amb97] (note that there as well as in [CGKS98] the use of interpolation is quite implicit). Those schemes improve upon [CGKS98] and attain communication complexity $O(n^{1/(2k-1)})$ for the case of k servers. The work of Ambainis uses recursion as well as polynomial interpolation. Other examples of PIR schemes of the first generation were given in [BIK05, Ito99, WY05] and [BIKR02, claims 3.1 and 3.2]. All those schemes are not recursive and have the same asymptotic parameters as the schemes of Ambainis [Amb97].

1.3.2 The second generation of LDCs and PIRs: recursion

The second generation of PIR schemes started with a breakthrough paper of Beimel et al. [BIKR02], who combined the earlier ideas of polynomial interpolation with a clever use of recursion to obtain k-server PIR schemes with $n^{O\left(\frac{\log\log k}{k \log k}\right)}$ asymptotic communication. The constructions of [BIKR02] gave improved upper bounds for communication complexity of k-server PIR for all values of $k \geq 3$. In particular [BIKR02] obtained 3-server PIR schemes with $O(n^{1/5.25})$ communication; 4-server PIR schemes with $O(n^{1/7.87})$ communication, and 5-server PIR schemes with $O(n^{1/10.83})$ communication.

Later, some of the results of [BIKR02] were given alternative proofs in the work of David Woodruff and the author [WY05] who also exploited an interplay between recursion and polynomial interpolation. ([WY05] matched the results of [BIKR02] for all $k \leq 26$, but did not prove the asymptotic bound.) The actual constructions of [BIKR02, WY05] are quite involved, and we do not include them in this thesis, but rather give a high-level overview of a specific 3-server PIR scheme of [WY05].

The scheme starts analogously to the PIR scheme from section 1.2. Given a database of length n, each server $\{\mathcal{S}_h\}_{h\in[3]}$ represents it by a homogeneous degree 7 polynomial D in $m = O(n^{1/7})$ variables over some finite field $\mathbb{F}_r$. The user $\mathcal{U}$ holds a point $W \in \mathbb{F}_r^m$ of Hamming weight 7 and wants to retrieve $D(W)$, while keeping the identity of W private. Rather than picking a random affine line containing the point W, and sending some of its points to servers, the user picks a random 2-dimensional affine plane π containing W, and sends each server $\mathcal{S}_h$ a line L_h in π. We assume the lines $\{L_h\}_{h\in[3]}$ are in general position. Note that now for every pair of servers $\mathcal{S}_{h_1}, \mathcal{S}_{h_2}$ there is a point $L_{h_1} \cap L_{h_2} \in \pi$ that is known to both of them. The user exploits this fact and for every pair of servers runs a separate 2-server PIR protocol to obtain some algebraic information about the restriction of D to π. Apart from that each server sends $\mathcal{U}$ the values and derivatives of the polynomial D at every point of its line. Finally the user combines all received information to obtain the restriction of D to π.

The results of [BIKR02] gave rise to the second generation of locally decodable codes. Those codes are obtained from PIR schemes and have smaller length than the codes of the first generation for all values of query complexity $k \geq 4$. In particular [BIKR02] obtained 4-query LDCs of length $\exp(n^{3/10})$; 5-query LDCs of length $\exp(n^{1/5})$, and 6-query LDCs of length $\exp(n^{1/7})$. For general query complexity k they achieved length $\exp\left(n^{O\left(\frac{\log\log k}{k \log k}\right)}\right)$.

1.3.3 The third generation of LDCs and PIRs: point removal

The third generation of locally decodable codes and private information retrieval schemes is the main contribution of this thesis. We introduce a novel (point removal) approach to constructing locally decodable codes and obtain vast improvements upon the earlier work. Our presentation is based on the key paper of the author [Yek07] and the follow up work of Kiran Kedlaya and the author [KY07].

Our results.

- Given a Mersenne number $m = 2^t - 1$ that has a large prime factor $p > m^{0.75}$, we design three query LDCs of length $N = \exp\left(n^{1/t}\right)$, for every n. Based on the current state of knowledge about Mersenne primes (i.e., primes of the form $2^t - 1$), this translates to a length of less than $\exp\left(n^{10^{-7}}\right)$, compared to $\exp\left(n^{1/2}\right)$ in the previous constructions. Our results for three query LDCs yield improvements of similar magnitude for larger values of query complexity k, via the generic reduction of [BIKR02].

- It has often been conjectured that there are infinitely many Mersenne primes. Under this conjecture, our constructions yield three query locally decodable codes of length $N = \exp\left(n^{O\left(\frac{1}{\log\log n}\right)}\right)$ *for infinitely many* n. Under a stronger (yet well accepted) conjecture [LPW, Pom80, Wag83] regarding the density of Mersenne primes, our constructions yield three query locally decodable codes of length $N = \exp\left(n^{O\left(\frac{1}{\log^{1-\epsilon}\log n}\right)}\right)$ *for all* n, for every $\epsilon > 0$.
- Our improvements in the parameters of locally decodable codes yield analogous improvements for private information retrieval schemes. We give 3-server PIR schemes with communication complexity of $O\left(n^{10^{-7}}\right)$ to access an n-bit database, compared to the previous best scheme with complexity $O(n^{1/5.25})$. Assuming again that there are infinitely many Mersenne primes, we get 3-server PIR schemes of communication complexity $n^{O(1/\log\log n)}$ *for infinitely many* n. Finally, assuming the conjecture regarding the density of Mersenne primes, we get 3-server PIR schemes of communication complexity $n^{O(1/\log^{1-\epsilon}\log n)}$ *for all* n, for every $\epsilon > 0$.

The results above were not expected by the community. After a decade of effort, many researchers in the area were pessimistic and believed that locally decodable codes with constant query complexity and subexponential length or private information retrieval schemes with constant number of servers and subpolynomial communication do not exist. In particular, such conjectures were published explicitly in [Gas04, section 9], [Gol05, conjecture 4.4].

Our technique. All previously known constructions of locally decodable codes and private information retrieval schemes are (explicitly or implicitly) centered around the idea of representing a message by an evaluation of a certain low degree polynomial over a finite field. Our constructions take a completely different approach.

We start by reducing the problem of constructing locally decodable codes to the problem of designing certain families of sets with restricted intersections. We next give a (basic) construction of such families that relies on linear algebra over finite fields. Our basic construction does not yield improved parameters for LDCs but has a simple geometric intuition underlining it: our universe is a high dimensional linear space (over a finite field), and our sets are lines and (unions of) affine hyperplanes. Our key insight that gives name to the method is that one can perform a surgery on the basic construction and greatly improve its parameters. Specifically, one can carefully *remove most of the points from lines* while preserving the right intersection properties.

The problem one needs to solve in order to successfully accomplish point removal is the following: one needs to design a set $S \subseteq \mathbb{F}_p^*$ for prime p that simultaneously satisfies two properties: (1) There exist two large sequences of vectors $u_1, \ldots, u_n$, $v_1, \ldots, v_n$ in some low dimensional space $\mathbb{F}_p^m$, such that the dot products $(u_i, v_i) = 0$ for all i, and the dot products $(u_j, v_i) \in S$ for all $i \neq j$. We refer to this property as the combinatorial niceness of S; (2) For a small integer k there exists a k sparse polynomial $\phi(x) \in \mathbb{F}_2[x]$ such that the common GCD of all polynomials of the form $\phi(x^\beta)$, $\beta \in S$ and the polynomial $x^p - 1$ is non-trivial. We refer to this property as the algebraic niceness of S.

Our construction of locally decodable codes thus comes in three steps: First we show that a set S exhibiting both combinatorial and algebraic niceness leads to good locally decodable codes. In particular the length n of the sequences $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ corresponds to the number of message bits we can encode, while the length of the codewords we build is $N = p^m$. So the longer the sequence and the smaller the dimension the better. The query complexity of our codes is given by the parameter k from the definition of algebraic niceness of S. This step of our construction is quite general and applies to vectors $u_1, \ldots, v_n$ and subsets S over any field. It leads us to the task of identifying good sets that are both combinatorially and algebraically nice, and these tasks narrow our choice of fields. As our second step we focus on combinatorial niceness. In general big sets tend to be "nicer" (allow longer sequences) than small ones. We show that every multiplicative subgroup of a prime field is combinatorially as nice as its cardinality would allow. This still leaves us with a variety of fields and subsets to work with. Finally as the last step we develop an insight into the algebraic niceness of sets. We focus on the very narrow case of primes p that are large prime factors of Mersenne numbers and the subgroup generated by the element 2 in $\mathbb{F}_p^*$. We manage to show that this subgroup is nice enough to get 3-query locally decodable codes, leading to our final result.

An alternative view of our constructions can be found in the follow up work of Raghavendra [Rag07].

Mersenne primes. As one can see above our results for locally decodable codes and private information retrieval schemes rely heavily on the known results regarding Mersenne numbers with large prime factors, (and in particular on the known results regarding Mersenne primes). To the best of author's knowledge our results are the first applications of Mersenne primes outside of number theory. We now briefly review the history of Mersenne primes starting from the ancient times. We also summarize the current knowledge regarding these numbers. Our exposition mostly follows [Pri].

Many ancient cultures were concerned with the relationship of a number with the sum of its divisors. Positive numbers that are equal to the sum of all of their positive divisors (excluding the number itself) were called *perfect*. Perfect numbers were often given mystic interpretations. The first four perfect numbers are $6 = 1 + 2 + 3$, $28 = 1 + 2 + 4 + 7 + 14$, 496, and 8128. By looking at prime factorization of these four numbers one can observe that each of them has the form $2^{t-1}(2^t - 1)$, where t is an integer and $2^t - 1$ is prime. This is not a coincidence. About 2300 years ago Euclid has proved that every number of such shape is perfect. (Much later Euler has proved a partial converse, namely that an *even* number is perfect if and only if it has the form $2^{t-1}(2^t - 1)$, where $2^t - 1$ is prime.)

The connection to perfect numbers mentioned above motivated mathematicians to look closer at prime numbers of the form $2^t - 1$. One of mathematicians who were interested in such primes was a French monk Marin Mersenne (1588-1648). In the preface to his book [Mer44] Mersenne stated that the numbers $2^t - 1$ were prime for

$$t = 2, 3, 5, 7, 13, 17, 19, 31, 67, 127 \text{ and } 257$$

and were composite for all other positive integers $t < 257$. Mersenne's conjecture was incorrect (he missed $61, 89, 107$ and included 67 and 257 which do not yield primes). Despite that nowadays integers of the form $2^t - 1$ are called Mersenne numbers, and primes of such form are called Mersenne primes.

Apart from the connection to perfect numbers, the other reasons Mersenne primes are so appealing are their succinctness and (since late 1870's) the existence of a very efficient deterministic Lucas-Lehmer [Ros88] primality test for integers of the form $2^t - 1$. The study of Mersenne primes that has started hundreds of years ago is still ongoing. A part of this activity is a search for large Mersenne primes. Nowadays this search involves using powerful modern computers. One of the notable contributors to the search is George Woltman. In 1996 he had the idea of using the internet to coordinate this search and developed the Great Internet Mersenne Prime Search project (GIMPS), that allows amateurs to join the search donating some computational power of their personal computers. GIMPS project has been quite successful: it has found a total of ten Mersenne primes, each of which was the largest known prime at the time of discovery. The largest known prime as of June 2007 is $2^{32,582,657} - 1$. This Mersenne prime was discovered on September 4, 2006 by Steven Boone and Curtis Cooper [CB].

Although the study of Mersenne primes has a very long history, so far mathematicians do not have answers even for the most basic questions regarding these numbers. In particular, it is not known whether the number of Mersenne primes is infinite. There is a feeling in the math community that it may take a long time before this question gets resolved [Sar]. A widely accepted conjecture states that (once found) the answer will be affirmative (i.e., there are infinitely many Mersenne primes). In fact, much stronger conjectures regarding the density of Mersenne primes have been made by Lenstra, Pomerance, and Wagstaff [LPW, Pom80, Wag83]. When presenting our results for locally decodable codes and private information retrieval schemes we give both unconditional results (based on the largest known Mersenne prime) and conditional results (under the assumption that the number of Mersenne primes is infinite).

1.3.4 Lower bounds

Katz and Trevisan [KT00] were the first to prove lower bounds for the length of locally decodable codes. Further work on lower bounds includes [GKST02, DJK$^+$02, Oba02, KdW04, WdW05, Woo07]. The length of optimal 2-query LDCs was settled by Kerenidis and de Wolf in [KdW04] and is $\exp(n)$. However for values of query complexity $k \geq 3$ we are still very far from closing the gap between lower and upper bounds. Specifically, the best lower bounds to date are of the form $\tilde{\Omega}\left(n^{1+1/(\lceil k/2 \rceil - 1)}\right)$ due to Woodruff [Woo07], while the best upper bounds are $\exp\left(n^{O_k(1/\log\log n)}\right)$ [Yek07] even under number theoretic conjectures.

The progress on lower bounds for private information retrieval has also been quite scarce. In what follows we list the known results for the two server case. The first nontrivial lower bound of $4 \log n$ is due to Mann [Man98]. Later it was improved to $4.4 \log n$ by Kerenidis and de Wolf [KdW04]. The current record of $5 \log n$ is due to Wehner and de Wolf [WdW05]. This leaves us with a tremendous gap to the best upper bound of $O(n^{1/3})$ [CGKS98]. It is quite interesting to note that this upper bound has never been improved since the initial paper of Chor et al. [CGKS98] in 1995. Although to date a number of different two server PIR schemes are known [CGKS98, BIK05, WY05] all of them have the same communication complexity.

Apart from the work on general lower bounds for PIR protocols, there has been some effort to establish (stronger) lower bounds for various restricted models of

PIR [Ito01, GKST02, BFG06, RY06]. In particular Itoh [Ito01] obtained polynomial lower bounds on communication complexity of one round PIR, under the assumption that each server returns a multilinear or affine function of its input. Goldreich et al. [GKST02] introduced the notion of *linear* PIR protocols, i.e., protocols where the servers are restricted to return linear combinations of the database bits to the user, and also the notion of *probe complexity,* i.e., the maximal number of bits the user needs to read from servers' answers in order to compute x_i. Goldreich et al. obtained polynomial lower bounds for communication complexity of two server linear PIR schemes whose probe complexity is constant. Later, their results were extended by Wehner and de Wolf [WdW05] who showed that the restriction of linearity can in fact be dropped.

Another restricted form of 2-server PIR was considered by Alexander Razborov and the author [RY06], who showed that every bilinear group-based PIR scheme requires $\Omega(n^{1/3})$ communication. A bilinear PIR scheme is a one round 2-server PIR scheme, where user computes the dot product of servers' responses to obtain the desired value of the i-th bit. A group based PIR scheme, is a PIR scheme, that involves servers representing database by a function on a certain finite group G, and allows user to retrieve the value of this function at any group element using the natural secret sharing scheme based on G. The model of bilinear group based PIR generalizes all PIR protocols known to date [RY06, appendix].

In chapter 4 of this thesis we present the results of [RY06] in full detail.

1.4 Applications of LDCs and PIRs

Earlier in this chapter we have talked about the application of locally decodable codes to data transmission and storage. We have also discussed the natural application of private information retrieval schemes. In this section we review some of the most notable other applications of LDCs and PIRs.

1.4.1 Secure multiparty computation

A fundamental result of Ben-Or, Goldwasser, and Wigderson [BOGW88] and Chaum, Crepeau, and Damgard [CCD88] from 1988 asserts that information-theoretic secure multiparty computation is feasible. Specifically, in [BOGW88, CCD88] it is shown that $k \geq 3$ players that are allowed to exchange messages over secure channels, can jointly compute any function of their local inputs while hiding the inputs from each other; i.e., one can always arrange a protocol as to ensure that after performing the joint computation any specific player gets no information about the inputs of other players (apart from the information contained in the value of the function).

In all known protocols for secure multiparty computation the communication complexity of the protocol grows linearly with the circuit size of the function being computed. This results in $\exp(n)$ communication for securely computing most of the functions of n bit inputs. A natural question that was explicitly asked in several papers from the late 1980's and early 1990's [DBR90, BFKR90] is whether *all* functions can be securely computed with only a polynomial (or at least a subexponential) amount of communication in the input length. It was observed by Ishai and Kushilevtiz [IK04] that this question is closely related to the complexity of private information retrieval schemes.

Our constructions of PIR schemes with subpolynomial amount of communication yield the first quantitative progress on the question mentioned above (via the reduction of [IK04]). Specifically, our results imply that a group of 18 or more players can securely compute any function of their n-bit inputs with a total communication of $\exp(n/\log^{1-\epsilon} n)$, for all n, for every $\epsilon > 0$, assuming the Lenstra, Pomerance, Wagstaff conjecture [LPW, Pom80, Wag83] regarding the density of Mersenne primes.

1.4.2 Average-case complexity

One of the earliest applications of locally decodable codes is the application to worst-case to average-case reductions. This application requires LDCs with polynomial length and polylogarithmic query complexity. Such codes are known to exist since

1990s [BFLS91, BFNW93, STV99] (in fact they predate the formal introduction of LDCs in [KT00]) and can be obtained by certain modifications [Tre04, section 3.4] of the classical Reed-Muller codes [MS77, vL82]. Our review of an example application of LDCs to worst-case to average case reductions mostly follows [Tre04, section 3.5].

Let L be an EXP-complete problem, and for an input length t let us consider the restriction of L to inputs of length t. We can see L restricted to these inputs as a binary string of length 2^t. Let us encode this string using a polynomial length locally decodable code C that has polylogarithmic query complexity and can tolerate some constant fraction of errors. We get a string of length $2^{O(t)} = 2^{t'}$, and let us think of this string as defining a new problem L' on inputs of length t'. If L was in EXP, then so is L'. The properties of the code C imply that a polynomial time algorithm for L' that is good on average (i.e., solves L' correctly on, say, some $1 - \epsilon$ fraction of the inputs in polynomial time) yields a probabilistic algorithm for L that works on all inputs, and EXP$\subseteq$BPP. This argument shows that if every problem in EXP can be solved well on average then EXP$\subseteq$BPP. A similar statement can be proved for PSPACE using a variant of this argument.

1.4.3 Other models of private information retrieval

A large number of extensions of the basic PIR model have been studied. These include extensions to t-private protocols, in which the user is protected against collusions of up to t servers [CGKS98, BIK05, BIW07]; extensions which protect the servers holding the database (in addition to the user), termed symmetric PIR [GIKM00, NP99]; and other extensions [BIM00, BS02, CIK$^+$01, DCIO01, GGM98, OS97]. In almost all extensions the best known solutions are obtained by adding some extra layers on top of a basic private information retrieval scheme. Therefore improving parameters of (basic) PIR schemes yields improvements for many other problems. For instance, see [BIW07] for a construction of improved t-private PIR schemes, based on PIR schemes from this thesis.

PIR was also studied in the *computational* setting where privacy should hold only against computationally bounded servers [KO97, Ste98, CMS99, BIKM99, DCMO00, KO00, KY01, Lip04, GR05, OS07]. In contrast to information-theoretic PIR, computational PIR protocols with sublinear communication exist even in the single-server case (under standard cryptographic assumptions). From a practical view point, single-server PIR protocols are preferable to multi-server ones for obvious reasons: they avoid the need to maintain replicated copies of the database or to compromise the user's privacy against several colluding servers. Moreover, single server protocols obtain better asymptotic communication complexity than information-theoretic protocols with a constant number of servers. However, for typical real-life parameters the known single-server protocols are less efficient than known multi-server (even 2-server) protocols. Furthermore, single-server protocols have some inherent limitations which can only be avoided in the multi-server setting. See [BIKR02, section 1] for further discussion.

1.5 Organization of the thesis

The main contribution of this thesis is a novel *point removal* approach to constructing locally decodable codes that yields vast improvements upon the earlier work. In chapter 2 we give a detailed treatment of the approach, and present our main results for LDCs. Chapter 3 deals with potential and limitations of the point-removal approach. We argue that further progress in the unconditional bounds via this method (under a fairly broad definition of the method) would imply progress on an old number theory question regarding the size of the largest prime factors of Mersenne numbers. Although chapters 2 and 3 are based on [Yek07, KY07], they contain some previously unpublished results. Specifically in the thesis we consider locally decodable codes over general (not necessarily binary) alphabets.

The last chapter 4 contains our results for private information retrieval schemes. We start by presenting tremendous improvements in upper bounds for PIR schemes involving three or more servers (that follow from improved upper bounds for LDCs). We then turn to the natural question regarding whether the two server PIR is truly intrinsically different. We argue that this may well be the case. We introduce a novel combinatorial approach to PIR and establish the optimality of the currently best known two server schemes a restricted although fairly broad model. The lower bounds part of chapter 4 is based on [RY06].

Chapter 2

Locally decodable codes via the point removal method

This chapter contains a detailed exposition of the point removal method for constructing LDCs. Our method can be broken into two parts. The first part is the reduction that shows how the existence of subsets of finite fields that simultaneously exhibit "nice" properties of two different kinds yields families of locally decodable codes with good parameters. The second part is the construction of "nice" subsets of finite fields.

Sections 2.1 and 2.2 of this chapter are preliminary. In section 2.3 we give a detailed treatment of the first part of our method for the narrow case of binary codes. We treat binary codes separately to have a simpler setup where we can (in an intuitive yet formal manner) demonstrate the combinatorial and geometric ideas that lie behind our method. While we believe that section 2.3 may be the most important part of the thesis (since it explains the intuition behind our approach), it can be skipped by the reader who is only interested in a succinct formal treatment of the constructions. After a detailed treatment of binary codes in section 2.3 we give a succinct treatment of general codes in section 2.4. As our main conclusion we identify the two "nice" properties of subsets of finite fields that (simultaneously) yield good codes. We call those properties combinatorial and algebraic niceness. The next two sections cover the second part of our method. In section 2.5 we construct combinatorially nice subsets of prime fields, and in section 2.6 we construct algebraically nice subsets of prime fields. Finally in section 2.7 we put the results of the previous sections together and summarize our improvements in upper bounds for locally decodable codes.

2.1 Notation

We use the following standard mathematical notation:

- $[s] = \{1, \ldots, s\}$;
- $\mathbb{Z}_n$ denotes integers modulo n;
- $\mathbb{F}_q$ is a finite field of q elements;
- $\mathbb{F}_q^*$ is the multiplicative group of $\mathbb{F}_q$;
- $d_H(x, y)$ denotes the Hamming distance between vectors x and y;
- (u, v) stands for the dot product of vectors u and v.
- For a linear space $L \subseteq \mathbb{F}_r^m$, $L^\perp$ denotes the *dual* space. That is,

$$L^\perp = \{u \in \mathbb{F}_r^m \mid \forall v \in L, (u, v) = 0\}.$$

2.2 Locally decodable codes

In this section we formally define locally decodable codes.

Definition 2.1 *An r-ary code $C : [r]^n \to [r]^N$ is said to be (k, δ, ϵ)-locally decodable if there exists a randomized decoding algorithm $\mathcal{A}$ such that*

1. *For all $x \in [r]^n$, $i \in [n]$ and $y \in [r]^N$ such that $d_H(C(x), y) \leq \delta N$: $Pr[\mathcal{A}^y(i) = x_i] \geq 1 - \epsilon$, where the probability is taken over the random coin tosses of the algorithm $\mathcal{A}$.*

2. $\mathcal{A}$ makes at most k queries to y.

In the special case when r is a prime power and the elements of the alphabet $[r]$ are in one to one correspondence with the elements of the finite field $\mathbb{F}_r$ it makes sense to talk about *linear* codes. A locally decodable code C is called linear if C is a linear transformation over $\mathbb{F}_r$. In this thesis we only consider codes over prime alphabets, and all our codes are linear.

2.3 Binary LDCs via point removal

In this section we give a detailed treatment of the first part of our method for the narrow case of binary codes. Our goal here is to explain the intuition behind the point removal approach, therefore we gradually build up our main construction, trying to provide the motivation for every choice that we make. Our final result is a claim that subsets of prime fields that exhibit certain properties (combinatorial and algebraic niceness) yield families of LDCs with very good parameters.

In section 2.3.1 we introduce certain combinatorial objects that we call regular intersecting families of sets. Those objects later serve as our tool to construct binary LDCs. In section 2.3.2 we present a linear algebraic construction of a regular intersecting family that yields locally decodable codes with good (although, not the best known) parameters. The notions of combinatorial and algebraic niceness of sets are used implicitly in this section. Our main construction in section 2.3.3 builds upon the construction of section 2.3.2 via the *point removal* procedure. We formally introduce combinatorial and algebraic niceness and show how the interplay between these two notions yields new LDCs.

2.3.1 Regular intersecting families of sets

Our constructions of locally decodable codes are linear. They are obtained by viewing the basis elements of the code and the decoding sets of the code as specifying a set system (where a vector corresponds to the set of coordinates on which it is non-zero), with some special intersection properties. We define these properties next. Let N, R and n be positive integers. Consider the set $[N]$. For $i \in [n]$, $r \in [R]$ let T_i and Q_{ir}, be subsets of $[N]$.

Definition 2.2 *We say that subsets T_i and Q_{ir} form a (k, n, N, R, s) regular intersecting family if the following conditions are satisfied:*

1. *k is odd;*

2. *For all $i \in [n]$, $|T_i| = s$;*

3. *For all $i \in [n]$ and $r \in [R]$, $|Q_{ir}| = k$;*

4. *For all $i \in [n]$ and $r \in [R]$, $Q_{ir} \subseteq T_i$;*

5. *For all $i \in [n]$ and $w \in T_i$, $|\{r \in [R] \mid w \in Q_{ir}\}| = (Rk)/s$, (i.e., T_i is uniformly covered by the sets Q_{ir});*

6. *For all $i, j \in [n]$ and $r \in [R]$ such that $i \neq j$, $|Q_{ir} \cap T_j| \equiv 0 \mod (2)$.*

The following proposition shows that regular intersecting families imply binary locally decodable codes.

Proposition 2.3 *A (k, n, N, R, s) regular intersecting family yields a binary linear code encoding n bits to N bits that is $(k, \delta, \delta Nk/s)$ locally decodable for all δ.*

Proof: For a set $S \subseteq [N]$ let $I(S) \in \{0,1\}^N$ denote its incidence vector. Formally, for $w \in [N]$ we set $I(S)_w = 1$ if $w \in S$; and $I(S)_w = 0$ otherwise. We define linear code C via its generator matrix $G \in \{0,1\}^{n \times N}$. For $i \in [n]$, we set the i-th row of G to be the incidence vector of the set T_i. Below is the description of the decoding algorithm $\mathcal{A}$. Given oracle access to y and input $i \in [n]$, the algorithm $\mathcal{A}$

1. picks $r \in [R]$ uniformly at random;

2. outputs the dot product $(y, I(Q_{ir}))$ over $\mathbb{F}_2$.

Note that since $|Q_{ir}| = k$, $\mathcal{A}$ needs only k queries into y to compute the dot product. It is easy to verify that the decoding is correct if $\mathcal{A}$ picks $r \in [R]$ such that all bits of xG in locations $h \in Q_{ir}$ are not corrupted:

$$(xG, I(Q_{ir})) = \sum_{j=1}^{n} x_j (I(T_j), I(Q_{ir})) = x_i (I(T_i), I(Q_{ir})) = x_i. \tag{2.1}$$

The second equality in formula (2.1) follows from part 6 of definition 2.2 and the last equality follows from parts 1,3 and 4 of definition 2.2. Now assume that up to δN bits of the encoding xG have been corrupted. Part 5 of definition 2.2 implies that there are at most $(\delta NRk)/s$ sets Q_{ir} that contain at least one corrupted location. Thus with probability at least $1 - (\delta Nk)/s$, the algorithm $\mathcal{A}$ outputs the correct value. ■

To the best of our knowledge regular intersecting families of sets have not been studied earlier. The closest combinatorial objects that have some literature are Ruzsa-Szemeredi (hyper)graphs [RS78, FLN$^+$02, SS05].

2.3.2 Basic construction

In this section we present our basic construction of regular intersecting families that yields binary k-query locally decodable codes of length $\exp\left(n^{1/(k-1)}\right)$ for prime values of $k \geq 3$. Note that for $k > 3$ the parameters that we get are inferior to the parameters of LDCs of the second generation (see section 1.3.2).

We choose our universe to be a high dimensional linear space over a prime field $\mathbb{F}_p$ and we choose sets T_i to be unions of cosets of certain hyperplanes and sets Q_{ir} to be lines. We argue the intersection properties based on elementary linear algebra. Let p be an odd prime and $m \geq p - 1$ be an integer.

Lemma 2.4 *Let* $n = \binom{m}{p-1}$. *There exist two families of vectors* $\{u_1, \ldots, u_n\}$ *and* $\{v_1, \ldots, v_n\}$ *in* $\mathbb{F}_p^m$, *such that*

- *For all* $i \in [n]$, $(u_i, v_i) = 0$;
- *For all* $i, j \in [n]$ *such that* $i \neq j$, $(u_j, v_i) \neq 0$.

Proof: Let $e \in \mathbb{F}_p^m$ be the vector that contains 1's in all the coordinates. We set vectors u_i to be incidence vectors of all possible $\binom{m}{p-1}$ subsets of $[m]$ of cardinality $(p-1)$. For every $i \in [n]$ we set $v_i = e - u_i$. It is straightforward to verify that this family satisfies the condition of the lemma. ■

Now we are ready to present our regular intersecting family. Set $N = p^m$ and $n = \binom{m}{p-1}$. Assume some bijection between the set $[N]$ and the space $\mathbb{F}_p^m$. For $i \in [n]$ set

$$T_i = \left\{ x \in \mathbb{F}_p^m \mid (u_i, x) \in \mathbb{F}_p^* \right\}.$$

Set $R = s = (p-1) \cdot p^{m-1}$. For each $i \in [n]$ assume some bijection between points of T_i and elements of $[R]$. For $i \in [n]$ and $r \in [R]$ let w_{ir} be the r-th point of T_i. Set

$$Q_{ir} = \{ w_{ir} + \lambda v_i \mid \lambda \in \mathbb{F}_p \}.[1]$$

Lemma 2.5 *For $i \in [n]$ and $r \in [R]$ sets T_i and Q_{ir} form a (p, n, N, R, s) regular intersecting family.*

Proof: We simply need to verify that all 6 conditions listed in definition 2.2 are satisfied.

1. Condition 1 is trivial.

2. Condition 2 is trivial.

3. Condition 3 is trivial.

4. Fix $i \in [n]$ and $r \in [R]$. Given that $(u_i, w_{ir}) \in \mathbb{F}_p^*$ let us show that $Q_{ir} \subseteq T_i$. By lemma 2.4 $(u_i, v_i) = 0$. Thus for every $\lambda \in \mathbb{F}_p : (u_i, w_{ir} + \lambda v_i) = (u_i, w_{ir})$. Condition 4 follows.

[1] Note that the sets Q_{ir} are not all distinct.

5. Fix $i \in [n]$ and $w \in T_i$. Note that

$$|\{r \in [R] \mid w \in Q_{ir}\}| = |\{w_{ir} \in T_i \mid \exists \lambda \in \mathbb{F}_p, w = w_{ir} + \lambda v_i\}| =$$

$$|\{w_{ir} \in T_i \mid \exists \lambda \in \mathbb{F}_p, w_{ir} = w - \lambda v_i\}| = p.$$

   It remains to notice that $Rp/s = p$. Condition 5 follows.

6. Fix $i, j \in [n]$ and $r \in [R]$ such that $i \neq j$. Note that

$$|Q_{ir} \cap T_j| = \left|\left\{\lambda \in \mathbb{F}_p \mid (u_j, w_{ir} + \lambda v_i) \in \mathbb{F}_p^*\right\}\right| =$$
$$\left|\left\{\lambda \in \mathbb{F}_p \mid ((u_j, w_{ir}) + \lambda(u_j, v_i)) \in \mathbb{F}_p^*\right\}\right| = p - 1.$$

   The last equality follows from the fact that $(u_j, v_i) \neq 0$, and therefore the univariate linear function $(u_j, w_{ir}) + \lambda(u_j, v_i)$ takes every value in $\mathbb{F}_p$ exactly once. It remains to notice that $p-1$ is even. Condition 6 follows.

■

Combining lemma 2.5 and proposition 2.3 we get

Corollary 2.6 *Let p be an odd prime and $m \geq p-1$ be an integer. There exists a binary linear code encoding $\binom{m}{p-1}$ bits to p^m bits that is $(p, \delta, \delta p^2/(p-1))$ locally decodable for all δ.*

It is now easy to convert the above result into a dense family (i.e., one that has a code for every message length n, as opposed to infinitely many n's) of p-query LDCs of length $\exp\left(n^{1/(p-1)}\right)$.

Theorem 2.7 *Let p be a fixed odd prime. For every positive integer n there exists a code of length* $\exp\left(n^{1/(p-1)}\right)$ *that is $(p, \delta, \delta p^2/(p-1))$ locally decodable for all δ.*

Proof: Given n, choose m to be the smallest integer such that $n \leq \binom{m}{p-1}$. Set $n' = \binom{m}{p-1}$. It is easy to verify that if n is sufficiently large we have $n' \leq 2n$. Given a message x of length n, we pad it with zeros to length n' and use the code from corollary 2.6 encoding x with a codeword of length $p^m = \exp\left(n^{1/(p-1)}\right)$. ■

2.3.3 Main construction: point removal

In the previous section we presented our basic linear algebraic construction of regular intersecting families. We chose sets T_i to be unions of cosets of certain hyperplanes. We chose sets Q_{ir} to be lines. The high-level idea behind our main construction, is to reduce the number of codeword locations queried by *removing some points from lines*; i.e., choosing sets Q_{ir} to be *proper subsets of lines* rather than whole lines.

Before we proceed to our main construction we introduce two central technical concepts of our method, namely *combinatorial* and *algebraic niceness.* We now give narrow definitions that are needed to construct binary codes via point removal method in linear spaces over prime fields. Later, (in section 2.4) we give more general definitions. Let p be an odd prime.

Definition 2.8 *A set $S \subseteq \mathbb{F}_p^*$ is called (m, n) combinatorially nice if there exist two families of vectors $\{u_1, \ldots, u_n\}$ and $\{v_1, \ldots, v_n\}$ in $\mathbb{F}_p^m$, such that*

- *For all $i \in [n]$, $(u_i, v_i) = 0$;*
- *For all $i, j \in [n]$ such that $i \neq j$, $(u_j, v_i) \in S$.*

Remark 2.9 Note that in lemma 2.4 we established that the set $S = \mathbb{F}_p^*$ is $\left(m, \binom{m}{p-1}\right)$ combinatorially nice for every integer $m \geq p - 1$.

Definition 2.10 *A set $S \subseteq \mathbb{F}_p^*$ is called k algebraically nice if k is odd and there exist two sets $S_0, S_1 \subseteq \mathbb{F}_p$ such that*

- *S_0 is not empty;*
- *$|S_1| = k$;*
- *For all $\alpha \in \mathbb{F}_p$ and $\beta \in S$: $|S_0 \cap (\alpha + \beta S_1)| \equiv 0 \mod (2)$.*

Remark 2.11 It is easy to verify that the set $S = \mathbb{F}_p^*$ is p algebraically nice. Simply pick $S_1 = \mathbb{F}_p$ and $S_0 = \mathbb{F}_p^*$.

The next lemma shows how an interplay between combinatorial and algebraic niceness yields regular intersecting families. It is the core of our construction.

Lemma 2.12 *Assume $S \subseteq \mathbb{F}_p^*$ is simultaneously (m, n) combinatorially nice and k algebraically nice. Let S_0 be the set from the definition of algebraic niceness of S. The set S yields a $(k, n, p^m, |S_0|p^{m-1}, |S_0|p^{m-1})$ regular intersecting family.*

Proof: For $i \in [n]$ let u_i, v_i be the vectors from the definition of combinatorial niceness. Set $N = p^m$ and $R = s = |S_0|p^{m-1}$. Assume a bijection between $[N]$ and $\mathbb{F}_p^m$. For all $i \in [n]$ set

$$T_i = \left\{x \in \mathbb{F}_p^m \mid (u_i, x) \in S_0\right\}.$$

For each $i \in [n]$ assume some bijection between $[R]$ and T_i. Let w_{ir} denote the r-th point of T_i. Set

$$Q_{ir} = \left\{w_{ir} + \lambda v_i \mid \lambda \in S_1\right\}.$$

It remains to verify that all 6 conditions listed in definition 2.2 are satisfied.

1. Condition 1 is trivial.

2. Condition 2 is trivial.

3. Condition 3 is trivial.

4. Fix $i \in [n]$ and $r \in [R]$. Given that $(u_i, w_{ir}) \in S_0$ let us show that $Q_{ir} \subseteq T_i$. Definition 2.8 implies that $(u_i, v_i) = 0$. Thus for every $\lambda \in S_1 : (u_i, w_{ir} + \lambda v_i) = (u_i, w_{ir})$. Condition 4 follows.

5. Fix $i \in [n]$ and $w \in T_i$. Note that

$$|\{r \in [R] \mid w \in Q_{ir}\}| = |\{w_{ir} \in T_i \mid \exists \lambda \in S_1, w = w_{ir} + \lambda v_i\}| =$$
$$|\{w_{ir} \in T_i \mid \exists \lambda \in S_1, w_{ir} = w - \lambda v_i\}| = |S_1| = k.$$

   It remains to notice that $Rk/s = k$. Condition 5 follows.

6. Fix $i, j \in [n]$ and $r \in [R]$ such that $i \neq j$. Note that

$$\begin{gathered}|Q_{ir} \cap T_j| = |\{\lambda \in S_1 \mid (u_j, w_{ir} + \lambda v_i) \in S_0\}| = \\ |\{\lambda \in S_1 \mid ((u_j, w_{ir}) + \lambda(u_j, v_i)) \in S_0\}| = \\ |S_0 \cap ((u_j, w_{ir}) + (u_j, v_i)S_1)| \equiv 0 \mod (2).\end{gathered}$$

The last equality follows from the fact that $(u_j, v_i) \in S$, and definition 2.10. Condition 6 follows.

■

Observe that one can derive a regular intersecting family with parameters from lemma 2.5 using lemma 2.12 in combination with remarks 2.9 and 2.11.

The next proposition that follows immediately by combining proposition 2.3 with lemma 2.12 is the heart of the first part of our construction of LDCs (for the case of binary codes).

Proposition 2.13 *Let p be an odd prime. Assume $S \subseteq \mathbb{F}_p^*$ is simultaneously (m, n) combinatorially nice and k algebraically nice. Let S_0 be the set from the definition of algebraic niceness of S. The set S yields a binary linear code encoding n bits to p^m bits that is $(k, \delta, \delta pk/|S_0|)$ locally decodable for all δ.*

Later we will see that for every Mersenne prime $p = 2^t - 1$ the multiplicative subgroup generated by the element 2 in $\mathbb{F}_p^*$ is three algebraically nice (lemma 2.30) and sufficiently combinatorially nice (lemma 2.21) to yield three query LDCs of length $\exp(n^{1/t})$ via the proposition above.

2.4 General LDCs via point removal

In this section we present a general treatment of the first part of our construction of locally decodable codes. We extend the results from the previous section in two ways: (1) we consider codes over alphabets $\mathbb{F}_r$, for arbitrary primes r, rather than only binary codes; (2) we consider nice subsets of arbitrary finite fields $\mathbb{F}_q$, rather than only prime fields. We start by defining combinatorial and algebraic niceness of subsets in the general setup, and then proceed to a succinct formal proof of the main propositions.

Definition 2.14 *Let q be a prime power. A set $S \subseteq \mathbb{F}_q^*$ is called (m, n) combinatorially nice if there exist two families of vectors $\{u_1, \ldots, u_n\}$ and $\{v_1, \ldots, v_n\}$ in $\mathbb{F}_q^m$, such that*

- *For all $i \in [n]$, $(u_i, v_i) = 0$;*
- *For all $i, j \in [n]$ such that $i \neq j$, $(u_j, v_i) \in S$.*

In many cases it will be more convenient for us to use the following definition of combinatorial niceness that involves a single parameter t.

Definition 2.15 *Let q be a prime power. A set $S \subseteq \mathbb{F}_q^*$ is called t combinatorially nice if for some $c > 0$ and every positive integer m, S is $(m, \lfloor cm^t \rfloor)$ combinatorially nice.*

Given a map f from a finite set to a field let $\mathrm{supp}(f)$ denote the *number* of elements of the set that are not mapped to zero. Now we proceed to the general definition of algebraic niceness.

Definition 2.16 *Let q be a prime power and r be a prime. A set $S \subseteq \mathbb{F}_q^*$ is called k algebraically over $\mathbb{F}_r$ nice if there exist two maps, $S_0 : \mathbb{F}_q \to \mathbb{F}_r$ and $S_1 : \mathbb{F}_q \to \mathbb{F}_r$ such that*

- $\mathrm{supp}(S_0) \neq 0$;
- $\mathrm{supp}(S_1) \leq k$;
- $\sum\limits_{\lambda \in \mathbb{F}_q} S_1(\lambda) \neq 0$;
- *For all $\alpha \in \mathbb{F}_q$ and $\beta \in S$:* $\sum\limits_{\lambda \in \mathbb{F}_q} S_0(\alpha + \beta\lambda) S_1(\lambda) = 0.$

We now proceed to our core lemma that shows how sets exhibiting both combinatorial and algebraic niceness yield locally decodable codes.

Lemma 2.17 *Let q be a prime power and r be a prime. Assume $S \subseteq \mathbb{F}_q^*$ is simultaneously (m,n) combinatorially nice, and k algebraically nice over $\mathbb{F}_r$. Let S_0 be the map from the definition of algebraic niceness of S. The set S yields an $\mathbb{F}_r$ linear code encoding n-long messages to q^m-long codewords that is $(k, \delta, \delta qk/\mathrm{supp}(S_0))$ locally decodable for all δ.*

Proof: Our proof comes in three steps. We specify encoding and local decoding procedures for our codes and then argue the lower bound for the probability of correct decoding. We use the notation from definitions 2.14 and 2.16.

Encoding: Our code will be linear. Therefore it suffices to specify the encoding of unit vectors $e_1, \ldots, e_n$, where e_j has length n and a unique non-zero coordinate j. We define the encoding of e_j to be a q^m long vector, whose coordinates are labelled by elements of $\mathbb{F}_q^m$. For all $w \in \mathbb{F}_q^m$ we set:

$$\mathrm{Enc}(e_j)_w = S_0\left((u_j, w)\right). \tag{2.2}$$

Local decoding: Given a (possibly corrupted) codeword y and an index $i \in [n]$, the decoding algorithm $\mathcal{A}$ picks $w \in \mathbb{F}_q^m$, such that $S_0((u_i, w)) \neq 0$ uniformly at random, reads $\mathrm{supp}(S_1) \leq k$ coordinates of y, and outputs the sum:

$$\frac{1}{S_0\left((u_i, w)\right) \sum\limits_{\lambda \in \mathbb{F}_q} S_1(\lambda)} \sum_{\lambda \in \mathbb{F}_q : S_1(\lambda) \neq 0} S_1(\lambda) y_{w + \lambda v_i}. \tag{2.3}$$

Probability of correct decoding: First we argue that decoding is always correct if $\mathcal{A}$ picks $w \in \mathbb{F}_q^m$ such that all coordinates of y with labels in the set $\{w + \lambda v_i\}_{\lambda : S_1(\lambda) \neq 0}$ are not corrupted. We need to show that for all $i \in [n]$, $x \in \mathbb{F}_r^n$ and $w \in \mathbb{F}_q^m$, such that $S_0((u_i, w)) \neq 0$:

$$\frac{1}{S_0\left((u_i, w)\right) \sum\limits_{\lambda \in \mathbb{F}_q} S_1(\lambda)} \sum_{\lambda \in \mathbb{F}_q : S_1(\lambda) \neq 0} S_1(\lambda) \left(\sum_{j=1}^{n} x_j \, \mathrm{Enc}(e_j) \right)_{w + \lambda v_i} = x_i. \tag{2.4}$$

To simplify the notation we put $c = 1/\left(S_0\left((u_i, w)\right) \sum\limits_{\lambda \in \mathbb{F}_q} S_1(\lambda)\right)$ and rewrite (2.4) as

$$c \sum_{\lambda \in \mathbb{F}_q} S_1(\lambda) \left(\sum_{j=1}^{n} x_j \ \mathrm{Enc}(e_j) \right)_{w+\lambda v_i} = x_i. \tag{2.5}$$

Note that

$$\begin{gathered} c \sum_{\lambda \in \mathbb{F}_q} S_1(\lambda) \left(\sum_{j=1}^{n} x_j \ \mathrm{Enc}(e_j) \right)_{w+\lambda v_i} = c \sum_{j=1}^{n} x_j \left(\sum_{\lambda \in \mathbb{F}_q} S_1(\lambda) \mathrm{Enc}(e_j)_{w+\lambda v_i} \right) = \\ c \sum_{j=1}^{n} x_j \left(\sum_{\lambda \in \mathbb{F}_q} S_1(\lambda) S_0((u_j, w + \lambda v_i)) \right). \end{gathered} \tag{2.6}$$

Now note that

$$\sum_{\lambda \in \mathbb{F}_q} S_1(\lambda) S_0((u_j, w + \lambda v_i)) = \sum_{\lambda \in \mathbb{F}_q} S_1(\lambda) S_0((u_j, w) + \lambda (u_j, v_i)) = \begin{cases} 1/c, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

The last identity above for $i = j$ follows from: $(u_i, v_i) = 0$, and the definition of constant c. The last identity for $i \neq j$ follows from $(u_j, v_i) \in S$ and the algebraic niceness of S. Combining (2.6) with the identity above we get (2.5).

Now assume that up to δ fraction of coordinates of y are corrupted. Let T_i denote the set of coordinates whose labels belong to $\left\{w \in \mathbb{F}_q^m \mid S_0((u_i, w)) \neq 0\right\}$. It is not hard to see that, $|T_i| = q^{m-1}\mathrm{supp}(S_0)$. Thus at most $\delta q/\mathrm{supp}(S_0)$ fraction of coordinates in T_i are corrupted. Let $Q_i = \left\{\{w + \lambda v_i\}_{\lambda \in \mathbb{F}_q : S_1(\lambda) \neq 0} \mid w : S_0((u_i, w)) \neq 0\right\}$ be the family of $\mathrm{supp}(S_1)$-tuples of coordinates that may be queried by $\mathcal{A}$. $(u_i, v_i) = 0$ implies that elements of Q_i uniformly cover the set T_i. Combining the last two observations we conclude that with probability at least $1 - \delta qk/\mathrm{supp}(S_0)$ $\mathcal{A}$ picks an uncorrupted $\mathrm{supp}(S_1) \leq k$ tuple and outputs the correct value of x_i. ■

The parameters of a locally decodable code that one gets by applying lemma 2.17 to a certain (nice) set S depend on support of S_0, where S_0 is a map from the definition of algebraic niceness of S. The next lemma shows that one can always ensure that the support of S_0 is large, and thus obtain a good dependence of the decoding error on the fraction of corrupted locations.

Lemma 2.18 *Let q be a prime power and r be a prime. Let $S \subseteq \mathbb{F}_q^*$ be a k algebraically nice set over $\mathbb{F}_r$. Let S_0, S_1 be maps from the definition of algebraic niceness of S. One can always redefine the map S_0 to satisfy* $\mathrm{supp}(S_0) \geq \lceil q(1-1/r) \rceil$.

Proof: Algebraic niceness of S implies that for all $\alpha \in \mathbb{F}_p$ and $\beta \in S$,

$$\sum_{\lambda \in \mathbb{F}_q} S_0(\alpha + \beta\lambda) S_1(\lambda) = 0.$$

Equivalently, for all $\alpha \in \mathbb{F}_p$ and $\beta \in S$,

$$\sum_{\lambda \in \mathbb{F}_q} S_0(\lambda) S_1((\lambda - \alpha)\beta^{-1}) = 0. \tag{2.7}$$

Our goal is to redefine the map S_0 to satisfy both (2.7) and $\mathrm{supp}(S_0) \geq \lceil q(1-1/r) \rceil$.

Consider a linear space $M = \mathbb{F}_r^q$ where coordinates of vectors are labelled by elements of $\mathbb{F}_q$. Note that there is a natural one to one correspondence between vectors in M and maps from $\mathbb{F}_q$ to $\mathbb{F}_r$. Specifically a map $f : \mathbb{F}_q \to \mathbb{F}_r$ corresponds to a vector $v \in M$ such that $v_\lambda = f(\lambda)$, for all $\lambda \in \mathbb{F}_q$.

Let $L \subseteq M$ be a linear subspace spanned by vectors corresponding to all maps $f(\lambda) = S_1((\lambda - \alpha)\beta^{-1})$, where $\alpha \in \mathbb{F}_q$ and $\beta \in S$. Observe that L is invariant under the actions of a 1-transitive permutation group (permuting the coordinates in accordance with addition in $\mathbb{F}_q$). This implies that the space $L^\perp$ is also invariant under the actions of the same group. Note that $L^\perp$ has positive dimension since it contains the vector corresponding to the map S_0. The last two observations imply that $L^\perp$ has a *full support,* i.e., for every $i \in [q]$ there exists a vector $v \in L^\perp$ such that $v_i \neq 0$. It is easy to verify that any linear subspace of $\mathbb{F}_r^q$ that has full support contains a vector of Hamming weight at least $\lceil q(1-1/r) \rceil$. Let $v \in L^\perp$ be such a vector. Redefining the map S_0 to be the map from $\mathbb{F}_q$ to $\mathbb{F}_r$ corresponding to vector v we conclude the proof. ■

The following propositions are the heart of the first part of our construction of LDCs. Combining lemmas 2.17 and 2.18 we get

Proposition 2.19 *Let q be a prime power and r be a prime. Assume $S \subseteq \mathbb{F}_q^*$ is simultaneously (m, n) combinatorially nice, and k algebraically nice over $\mathbb{F}_r$. The set S yields an $\mathbb{F}_r$ linear code encoding n-long messages to q^m-long codewords that is $(k, \delta, \delta kr/(r-1))$ locally decodable for all δ.*

Using proposition 2.19 in combination with a single parameter definition of combinatorial niceness we get

Proposition 2.20 *Let q be a prime power and r be a prime. Assume $S \subseteq \mathbb{F}_q^*$ is simultaneously t combinatorially nice, and k algebraically nice over $\mathbb{F}_r$; then for every $n > 0$ there exists an $\mathbb{F}_r$ linear code encoding n-long messages to $\exp(n^{1/t})$-long codewords that is $(k, \delta, \delta kr/(r-1))$ locally decodable for all δ.*

Proof: Let $c > 0$ be the constant from the (single parameter) definition of combinatorial niceness of S. Given a message of length n we pad it with zeros to get a message of length n', where $n' \geq n$ is the smallest integer of the form $\lfloor cm^t \rfloor$; and then use the code from proposition 2.19. It is not hard to verify that the padding results in a most a constant (multiplicative) blow-up in the message length, and thus the length of our code is $\exp(n^{1/t})$. ■

Propositions 2.19 and 2.20 identify two properties of subsets of finite fields that together yield good locally decodable codes. These properties are combinatorial and algebraic niceness. Our next goal is to construct nice subsets. In the next sections we show that if primes p and r are such that p is a large factor of $r^t - 1$; then the multiplicative subgroup generated by the number r in $\mathbb{F}_p^*$ is sufficiently (algebraically and combinatorially) nice to yield constant-query LDCs of length $\exp(n^{1/t})$ over $\mathbb{F}_r$ for all message lengths n.

2.5 Combinatorially nice subsets of $\mathbb{F}_p^*$

In this section we study combinatorial niceness and show that multiplicative subgroups of prime fields are combinatorially nice.

For $w \in \mathbb{F}_p^m$ and a positive integer l, let $w^{\otimes l} \in \mathbb{F}_p^{m^l}$ denote the l-th tensor power of w. Coordinates of $w^{\otimes l}$ are labelled by all possible sequences in $[m]^l$ and $w^{\otimes l}_{i_1,\ldots,i_l} = \prod\limits_{j=1}^{l} w_{i_j}$. Our next goal is to establish the following

Lemma 2.21 *Let p be a prime and $m \geq p-1$ be an integer. Suppose S is a subgroup of $\mathbb{F}_p^*$; then S is $\left(\binom{m-1+(p-1)/|S|}{(p-1)/|S|}, \binom{m}{p-1}\right)$ combinatorially nice.*

Proof: Let $n = \binom{m}{p-1}$. For $i \in [n]$ let vectors u_i'' and v_i'' in $\mathbb{F}_p^m$ be the same as vectors u_i, v_i in the proof of lemma 2.4, i.e., vectors u_i'' are incidence vectors of all possible subsets of $[m]$ of cardinality $(p-1)$ and vectors v_i'' are their complements. Recall that

- For all $i \in [n]$, $(u_i'', v_i'') = 0$;
- For all $i, j \in [n]$ such that $i \neq j$, $(u_j'', v_i'') \neq 0$.

Let l be a positive integer and u, v be vectors in $\mathbb{F}_p^m$. Observe that

$$\begin{aligned}(u^{\otimes l}, v^{\otimes l}) = \sum_{(i_1,\ldots,i_l)\in[m]^l} \left(\prod_{j=1}^{l} u_{i_j} \prod_{j=1}^{l} v_{i_j}\right) = \\ \sum_{(i_1,\ldots,i_l)\in[m]^l} \left(\prod_{j=1}^{l} u_{i_j} v_{i_j}\right) = \left(\sum_{i_1\in[m]} u_{i_1} v_{i_1}\right) \ldots \left(\sum_{i_l\in[m]} u_{i_l} v_{i_l}\right) = (u, v)^l.\end{aligned} \tag{2.8}$$

Let $l = (p-1)/|S|$. For $i \in [n]$ set $u_i' = u_i''^{\otimes l}$ and $v_i' = v_i''^{\otimes l}$. Formula (2.8) and cyclicity of $\mathbb{F}_p^*$ yield

- For all $i \in [n]$, $(u_i', v_i') = 0$;
- For all $i, j \in [n]$ such that $i \neq j$, $(u_j', v_i') \in S$.

Note that vectors u_i' and v_i' are $m^{(p-1)/|S|}$ long. Therefore at this point we have already shown that the set S is $\left(m^{(p-1)/|S|}, \binom{m}{p-1}\right)$ combinatorially nice.

Let w be an arbitrary vector in $\mathbb{F}_p^m$. Note that the value of $w^{\otimes l}_{i_1,\ldots,i_l}$ depends on the *multi-set* $\{i_1, \ldots, i_l\}$ rather than the sequence $i_1, \ldots, i_l$. Thus many coordinates of $w^{\otimes l}$ contain identical (and therefore redundant) values. We are going to reduce the length of vectors u_i' and v_i' using this observation. Let $F(m, l)$ denote the family of all multi-subsets of $[m]$ of cardinality l. Note that $|F(m, l)| = \binom{m-1+l}{l}$. For a multi-set $\sigma \in F(m, l)$ let $c(\sigma)$ denote the number of sequences in $[m]^l$ that represent σ. Now we are ready to define vectors u_i and v_i in $\mathbb{F}_p^{|F(m,l)|}$. The Coordinates of the vectors u_i and v_i are labelled by multi-sets $\sigma \in F(m, l)$. For all $i \in [n]$ and $\sigma \in F(m, l)$ we set

$$(u_i)_\sigma = c(\sigma)(u_i')_\sigma \quad \text{and} \quad (v_i)_\sigma = (v_i')_\sigma.$$

It is easy to verify that for all $i, j \in [n]$, $(u_j, v_i) = \left(u_j', v_i'\right)$. Combining this observation with the properties of vectors u_i' and v_i' that were established earlier, we conclude that the set S is $\left(\binom{m-1+(p-1)/|S|}{(p-1)/|S|}, \binom{m}{p-1}\right)$ combinatorially nice. ■

We now give a simple corollary to lemma 2.21 that uses a single parameter definition of combinatorial niceness.

Lemma 2.22 *Let p be a prime. Suppose S is a subgroup of $\mathbb{F}_p^*$; then S is $|S|$ combinatorially nice.*

Proof: Let $t = |S|$. We need to specify a constant $c > 0$ such that for every positive integer m there exist two $n = \lfloor cm^t \rfloor$-sized collections of m long vectors over $\mathbb{F}_p$ satisfying:

- For all $i \in [n]$, $(u_i, v_i) = 0$;
- For all $i, j \in [n]$ such that $i \neq j$, $(u_j, v_i) \in S$.

First assume that m has the shape $m = \binom{m'-1+(p-1)/t}{(p-1)/t}$, for some integer $m' \geq p - 1$. In this case lemma 2.21 gives us a collection of $n = \binom{m'}{p-1}$ vectors with the right properties. Observe that $n \geq cm^t$ for a constant c that depends only on p and t. Now assume m does not have the right shape, and let m_1 be the largest integer smaller than m that does have it. In order to get vectors of length m we use vectors of length m_1 coming from lemma 2.21 padded with zeros. It is not hard to verify that such a construction still gives us $n \geq cm^t$ large families of vectors for a suitably chosen constant c. ■

2.6 Algebraically nice subsets of $\mathbb{F}_p^*$

In the previous section we studied combinatorial niceness and established that multiplicative subgroups of prime fields are combinatorially nice. In this section we study algebraic niceness, and show that (under certain constraints on p and r) the multiplicative subgroup generated by r in $\mathbb{F}_p^*$ is algebraically nice over $\mathbb{F}_r$.

We start by introducing some notation. Let p and r be distinct primes.

- $\mathrm{ord}_p(r)$ denotes the smallest integer t such that $p \mid r^t - 1$.
- $\langle r \rangle \subseteq \mathbb{F}_p^*$ denotes the multiplicative subgroup of $\mathbb{F}_p^*$ generated by r. Clearly, $|\langle r \rangle| = \mathrm{ord}_p(r)$;
- $\overline{\mathbb{F}}$ denotes the algebraic closure of the field $\mathbb{F}$;
- $C_r^p \subseteq \overline{\mathbb{F}}_r^*$ denotes the multiplicative subgroup of p-th roots of unity in $\overline{\mathbb{F}}_r$.

Definition 2.23 *Let p and r be distinct primes. We say that there is a nontrivial k dependence between the elements of C_r^p if there exist $\zeta_1, \ldots, \zeta_k \in C_r^p$ and $\sigma_1, \ldots, \sigma_k \in \mathbb{F}_r$ such that*

$$\sigma_1\zeta_1 + \ldots + \sigma_k\zeta_k = 0 \quad \textit{and} \quad \sigma_1 + \ldots + \sigma_k \neq 0. \tag{2.9}$$

Lemma 2.24 *Let p and r be distinct primes. Suppose there exists a nontrivial k dependence between the elements of C_r^p; then $\langle r \rangle \subseteq \mathbb{F}_p^*$ is k algebraically nice over $\mathbb{F}_r$.*

Proof: In what follows we define a map $S_1 : \mathbb{F}_p \to \mathbb{F}_r$ and prove the existence of a map $S_0 : \mathbb{F}_p \to \mathbb{F}_r$ such that together S_0 and S_1 yield k algebraic niceness of $\langle r \rangle$

over $\mathbb{F}_r$. Identity (2.9) implies that for some $k' \leq k$ there exist k' *distinct* p-th roots of unity $\zeta_1, \ldots, \zeta_{k'} \in C_r^p$ such that for some $\sigma_1, \ldots, \sigma_{k'} \in \mathbb{F}_r$

$$\sigma_1\zeta_1 + \ldots + \sigma_{k'}\zeta_{k'} = 0 \text{ and } \sigma_1 + \ldots + \sigma_{k'} \neq 0. \tag{2.10}$$

Let $t = \operatorname{ord}_p(r)$. Observe that $C_r^p \subseteq \mathbb{F}_{r^t}$. Let g be a generator of C_r^p. Identity (2.10) yields $\sigma_1 g^{\gamma_1} + \ldots + \sigma_{k'} g^{\gamma_{k'}} = 0$, for some distinct values $\{\gamma_i\}_{i\in[k']}$ in $\mathbb{Z}_p$. We define $S_1(\gamma_i) = \sigma_i$, for all $i \in [k']$; and $S_1(\lambda) = 0$ for all other $\lambda \in \mathbb{F}_p$. Identity (2.10) yields $\operatorname{supp}(S_1) \leq k$ and $\sum\limits_{\lambda\in\mathbb{F}_p} S_1(\lambda) \neq 0$.

Now our goal is to prove the existence of a (nonzero) map $S_0 : \mathrm{F}_p \to \mathbb{F}_r$ such that for all $\alpha \in \mathbb{F}_p$ and $\beta \in S$:

$$\sum_{\lambda\in\mathbb{F}_p} S_0(\alpha + \beta\lambda)S_1(\lambda) = 0.$$

Equivalently, we need (a nonzero) map S_0 such that for all $\alpha \in \mathbb{F}_p$ and $\beta \in S$:

$$\sum_{\lambda\in\mathbb{F}_p} S_0(\lambda)S_1((\lambda - \alpha)\beta^{-1}) = 0. \tag{2.11}$$

Consider a natural one to one correspondence between maps $S' : \mathbb{F}_p \to \mathbb{F}_r$ and polynomials $\phi_{S'}(x)$ in the ring $\mathbb{F}_r[x]/(x^p - 1)$:

$$\phi_{S'}(x) = \sum_{\lambda\in\mathbb{Z}_p} S'(\lambda)x^\lambda.$$

Clearly, for every map $S' : \mathbb{F}_p \to \mathbb{F}_r$ and every fixed $\alpha, \beta \in \mathbb{F}_p$, such that $\beta \neq 0$:

$$\phi_{S'((\lambda-\alpha)\beta^{-1})}(x) = \sum_{\lambda\in\mathbb{F}_p} S'((\lambda - \alpha)\beta^{-1})x^\lambda = \sum_{\lambda\in\mathbb{F}_p} S'(\lambda)x^{\alpha+\beta\lambda} = x^\alpha\phi_{S'}(x^\beta).$$

Let α be a variable ranging over $\mathbb{F}_p$ and β be a variable ranging over $\langle r\rangle$. We are going to argue the existence of a map $S_0 : \mathbb{F}_p \to \mathbb{F}_r$ that has satisfies (2.11), by showing that all polynomials $\phi_{S_1((\lambda-\alpha)\beta^{-1})}$ belong to a certain linear space $L \in \mathbb{F}_r[x]/(x^p - 1)$ of dimension less than p. In this case any (nonzero) map $T : \mathbb{F}_p \to \mathbb{F}_r$ such that $\phi_T \in L^{\perp}$ can be used as the map S_0. Let $\tau(x) = \gcd(x^p - 1, \phi_{S_1}(x))$. Note that $\tau(x) \neq 1$ since g is a common root of $x^p - 1$ and $\phi_{S_1}(x)$. Let L be the space of polynomials in $\mathbb{F}_r[x]/(x^p - 1)$ that are multiples of $\tau(x)$. Clearly, $\dim L = p - \deg \tau$. Fix some $\alpha \in \mathbb{F}_p$ and $\beta \in \langle r \rangle$. Let us prove that $\phi_{S_1((\lambda-\alpha)\beta^{-1})}(x)$ is in L :

$$\phi_{S_1((\lambda-\alpha)\beta^{-1})}(x) = x^{\alpha}\phi_{S_1}(x^{\beta}) = x^{\alpha}(\phi_{S_1}(x))^{\beta}.$$

The last identity above follows from the fact that for any $f \in \mathbb{F}_r[x]$ and any positive integer $i : f(x^{r^i}) = (f(x))^{r^i}$. ■

Lemma 2.24 reduces the task of proving k algebraic niceness of $\langle r \rangle \subseteq \mathbb{F}_p^*$ to certifying the existence of a nontrivial k dependence in C_r^p. In the following subsections we present several sufficient conditions for the existence of such a dependence.

Our first sufficient condition (lemma 2.25) is the following: p is a Mersenne prime and $r = 2$. The proof that this condition suffices is simple and self-contained. This result alone already yields most of our improvements for binary locally decodable codes (see lemma 2.30 and section 2.7.1). Two weaker sufficient conditions are given in lemmas 2.26 and 2.29. The proofs of those lemmas are quite technical. Those lemmas are later used to obtain the most general form of our results for LDCs (see section 2.7.2).

2.6.1 3-dependencies between p-th roots: sufficient conditions

Lemma 2.25 *Suppose $p = 2^t - 1$ is a Mersenne prime; then there exists a nontrivial three dependence in C_2^p.*

Proof: Observe that the polynomial $x^p - 1 = x^{2^t-1} - 1$ splits into distinct linear factors in the finite field $\mathbb{F}_{2^t}$. Therefore $C_2^p = \mathbb{F}_{2^t}^*$. Pick $\zeta_1 \neq \zeta_2$ in C_2^p arbitrarily. Set $\zeta_3 = \zeta_1 + \zeta_2$. Note that $\zeta_3 \in C_2^p$ and $\zeta_1 + \zeta_2 + \zeta_3 = 0$. ■

Now we generalize lemma 2.25 and show that a substantially weaker condition on p and r is still sufficient. Our argument relies on the classical Weil bound [LN83, p. 330] for the number of rational points on curves over finite fields.

Lemma 2.26 *Let p and r be distinct primes. Suppose $\mathrm{ord}_p(r) < (4/3)\log_r p$; then there exists a nontrivial three dependence in C_r^p.*

Proof: We start with a brief review of some basic concepts of projective algebraic geometry [CLO96]. Let $\mathbb{F}$ be a field, and $f \in \mathbb{F}[x,y,z]$ be a homogeneous polynomial. A triple $(x_0, y_0, z_0) \in \mathbb{F}^3$ is called a zero of f if $f(x_0, y_0, z_0) = 0$. A zero is called nontrivial if it is different from the origin. An equation $f = 0$ defines a projective plane curve χ_f. Nontrivial zeros of f considered up to multiplication by a scalars are called $\mathbb{F}$-rational points of χ_f. If $\mathbb{F}$ is a finite field it makes sense to talk about the number of $\mathbb{F}$-rational points on a curve.

Let $t = \mathrm{ord}_p(r)$. Note that $C_r^p \subseteq \mathbb{F}_{r^t}$. Pick $\{\sigma_i\}_{i\in[3]}$ in $\mathbb{F}_r^*$ such that $\sigma_1+\sigma_2+\sigma_3 \neq 0$. Consider a projective plane curve χ defined by

$$\sigma_1 x^{(r^t-1)/p} + \sigma_2 y^{(r^t-1)/p} + \sigma_3 z^{(r^t-1)/p} = 0. \tag{2.12}$$

Let us call a point a on χ trivial if one of the coordinates of a is zero. Clearly, there are at most $3(r^t-1)/p$ trivial points on χ. Note that every nontrivial $\mathbb{F}_{r^t}$-rational point of χ yields a nontrivial 3-dependence in C_r^p (since $\mathbb{F}_{r^t}^*$ is cyclic). The classical Weil bound [LN83, p. 330] provides an estimate

$$|N_q - (q+1)| \leq (d-1)(d-2)\sqrt{q} \tag{2.13}$$

for the number N_q of $\mathbb{F}_q$-rational points on an arbitrary smooth projective plane curve of degree d. (2.13) implies that in case

$$r^t + 1 > \left(\frac{r^t-1}{p} - 1\right)\left(\frac{r^t-1}{p} - 2\right) r^{t/2} + 3\frac{r^t-1}{p} \tag{2.14}$$

there exists a nontrivial point on the curve (2.12). Note that (2.14) follows from

$$r^t + 1 > \left(\frac{r^t}{p}\right)\left(\frac{r^t}{p}\right) r^{t/2} - \frac{2r^{3t/2}}{p} + \frac{3r^t}{p}, \tag{2.15}$$

and (2.15) follows from

$$r^t > r^{2t+t/2}/p^2 \quad \text{and} \quad 2r^{t/2} > 3.$$

Now note that the first inequality above follows from $t < (4/3)\log_r p$. To prove the second inequality observe that $r \geq 3$ implies $2r^{1/2} > 3$, and $r = 2$ implies $t \geq 2$. ■

2.6.2 k-dependencies between p-th roots: a sufficient condition

In this section we show that one can further relax the conditions of lemma 2.26 and still ensure the existence of nontrivial k dependencies in C_r^p, (for $k \geq 3$). Our proof is quite technical and comes in three steps. First we briefly review the notion of (additive) Fourier coefficients of subsets of $\mathbb{F}_{r^t}$. Next, we invoke a folklore argument to show that subsets of $\mathbb{F}_{r^t}$ with appropriately small nontrivial Fourier coefficients contain nontrivial k dependecies. Finally, we use a recent result of Bourgain and Chang [BC06] (generalizing the classical estimate for Gauss sums) to argue that (under certain constraints on p and r) all nontrivial Fourier coefficients of C_r^p are small.

For a prime r let $\mathbb{C}_r$ denote the multiplicative group of complex r-th roots of unity. Let $e \in \mathbb{C}_r$ be an r-th root other than the identity. For $x \in \mathbb{F}_{r^t}$ let $\mathrm{Tr}(x) = x + x^r + \ldots + x^{r^{t-1}}$ denote the trace of x. It is not hard to verify that for all x, $\mathrm{Tr}(x) \in \mathbb{F}_r$. Characters of $\mathbb{F}_{r^t}$ are homomorphisms from the additive group of $\mathbb{F}_{r^t}$ into $\mathbb{C}_r$. There exist r^t characters. We denote characters by χ_a, where a ranges in $\mathbb{F}_{r^t}$, and set $\chi_a(x) = e^{\mathrm{Tr}(ax)}$. Let $C(x)$ denote the incidence function of a set $C \subseteq \mathbb{F}_{r^t}$. For arbitrary $a \in \mathbb{F}_{r^t}$ the Fourier coefficient $\hat{C}(\chi_a)$ is defined by $\hat{C}(\chi_a) = \sum \chi_a(x)C(x)$, where the sum is over all $x \in \mathbb{F}_{r^t}$. Fourier coefficient $\hat{C}(\chi_0) = |C|$ is called trivial, and other Fourier coefficients are called nontrivial. In what follows $\sum_a$ stands for summation over all r^t elements of $\mathbb{F}_{r^t}$. We need the following two standard properties of characters and Fourier coefficients.

$$\sum_a \chi_a(x) = \begin{cases} r^t, & \text{if } x = 0; \\ 0, & \text{otherwise.} \end{cases} \tag{2.16}$$

$$\sum_a \left|\hat{C}(\chi_a)\right|^2 = r^t|C|. \tag{2.17}$$

The following lemma is a folklore.

Lemma 2.27 *Let $C \subseteq \mathbb{F}_{r^t}$ and $k \geq 3$ be an integer such that there exist $\{\sigma_i\}_{i\in[k]}$ in $\mathbb{F}_r^*$, where $\sum_{i\in[k]} \sigma_i \neq 0$. Let F be the largest absolute value of a nontrivial Fourier coefficient of C. Suppose*

$$\frac{F}{|C|} < \left(\frac{|C|}{r^t}\right)^{1/(k-2)} \tag{2.18}$$

then there exists a nontrivial k dependence between the elements of C.

Proof: Let $M(C) = \#\{\zeta_1, \ldots, \zeta_k \in C \mid \sigma_1\zeta_1 + \ldots + \sigma_k\zeta_k = 0\}$. Identity (2.16) yields

$$M(C) = \frac{1}{r^t} \sum_{x_1,\ldots,x_k \in \mathbb{F}_{r^t}} C(x_1)\ldots C(x_k) \sum_a \chi_a(\sigma_1 x_1 + \ldots + \sigma_k x_k). \tag{2.19}$$

Note that $\chi_a(\sigma_1 x_1 + \ldots + \sigma_k x_k) = \chi_{\sigma_1 a}(x_1)\ldots\chi_{\sigma_k a}(x_k)$. Changing the order of summation in (2.19) we get

$$M(C) = \frac{1}{r^t} \sum_a \sum_{x_1,\ldots,x_k \in \mathbb{F}_{r^t}} C(x_1)\ldots C(x_k)\chi_{\sigma_1 a}(x_1)\ldots\chi_{\sigma_k a}(x_k) \tag{2.20}$$

Separating the term corresponding to $a = 0$ in the right hand side of (2.20) we get

$$M(C) = \frac{|C|^k}{r^t} + \frac{1}{r^t}\sum_{a\neq 0}\prod_{i=1}^{k} \hat{C}(\chi_{\sigma_i a}) \geq \frac{|C|^k}{r^t} - \frac{1}{r^t}\sum_{a\neq 0}\prod_{i=1}^{k} \left|\hat{C}(\chi_{\sigma_i a})\right|. \tag{2.21}$$

Using the generalized Holder's inequality [BB65, p. 20] we obtain

$$\sum_{a\neq 0}\prod_{i=1}^{k}\left|\hat{C}(\chi_{\sigma_i a})\right| \leq \prod_{i=1}^{k}\left(\sum_{a\neq 0}\left|\hat{C}(\chi_{\sigma_i a})\right|^k\right)^{1/k}. \tag{2.22}$$

Note that for every $i \in [k]$ we have

$$\sum_{a\neq 0}\left|\hat{C}(\chi_{\sigma_i a})\right|^k \leq F^{k-2}\sum_{a}\left|\hat{C}(\chi_{\sigma_i a})\right|^2 = F^{k-2}r^t|C|, \tag{2.23}$$

where the last identity follows from (2.17). Combining (2.21), (2.22) and (2.23) we get

$$M(C) \geq \frac{|C|^k}{r^t} - F^{k-2}|C|, \tag{2.24}$$

and conclude that we conclude that (2.18) implies $M(C) > 0$. ■

The following lemma is due to Bourgain and Chang [BC06, theorem 1].

Lemma 2.28 *Assume that $n \mid r^t - 1$ and satisfies the condition*

$$\gcd\left(n, \frac{r^t-1}{r^{t'}-1}\right) < r^{t(1-\epsilon)-t'}, \quad \textit{for all} \quad 1 \leq t' < t,\ t' \mid t,$$

where $\epsilon > 0$ is arbitrary and fixed. Then for all $a \in \mathbb{F}_{r^t}^*$

$$\left|\sum_{x\in\mathbb{F}_{r^t}} e^{\mathrm{Tr}(ax^n)}\right| < c_1 r^{t(1-\delta)}, \tag{2.25}$$

where $\delta = \delta(\epsilon) > 0$ and $c_1 = c_1(\epsilon)$ are constants.

Below is the main result of this subsection. Recall that C_r^p denotes the set of p-th roots of unity in $\overline{\mathbb{F}}_r$.

Lemma 2.29 *For every $c > 0$ and prime r there exists an integer $k = k(c, r)$ such that the following implication holds. If $p \neq r$ is a prime and $\mathrm{ord}_p(r) < c\log_r p$ then there is a nontrivial k dependence between the elements of C_r^p.*

Proof: Note that the sum of all p-th roots of unity in $\overline{\mathbb{F}}_r$ is zero. Therefore given r and c it suffices to prove the existence of $k = k(c,r)$ that works for all *sufficiently large* p. Let $t = \mathrm{ord}_p(r)$. Observe that $p > r^{t/c}$. Assume p is sufficiently large so that $t > 2c$. Next we show that the precondition of lemma 2.28 holds for $n = (r^t - 1)/p$ and $\epsilon = 1/(2c)$. Let $t' \mid t$ and $1 \leq t' < t$. Clearly $\gcd(r^{t'} - 1, p) = 1$. Therefore

$$\gcd\left(\frac{r^t - 1}{p}, \frac{r^t - 1}{r^{t'} - 1}\right) = \frac{r^t - 1}{p(r^{t'} - 1)} < \frac{r^{t(1-1/c)}}{r^{t'} - 1}, \tag{2.26}$$

where the inequality follows from $p > r^{t/c}$. Clearly, $t > 2c$ yields $r^{t/(2c)}/2 > 1$. Multiplying the right hand side of (2.26) by $r^{t/(2c)}/2$ and using $2(r^{t'} - 1) \geq r^{t'}$ we get

$$\gcd\left(\frac{r^t - 1}{p}, \frac{r^t - 1}{r^{t'} - 1}\right) < r^{t(1-1/(2c))-t'}. \tag{2.27}$$

Combining (2.27) with lemma 2.28 we conclude that there exist $\delta > 0$ and c_1 such that for all $a \in \mathbb{F}_{r^t}^*$

$$\left| \sum_{x \in \mathbb{F}_{r^t}} e^{\mathrm{Tr}\left(ax^{(r^t-1)/p}\right)} \right| < c_1 r^{t(1-\delta)}. \tag{2.28}$$

Observe that $x^{(r^t-1)/p}$ takes every value in C_r^p exactly $(r^t - 1)/p$ times when x ranges over $\mathbb{F}_{r^t}^*$. Thus (2.28) implies

$$(r^t - 1)(F/p) < c_1 r^{t(1-\delta)} + 1, \tag{2.29}$$

where F denotes the largest absolute value of a nontrivial Fourier coefficient of C_r^p. Assuming that t is sufficiently large, we get

$$(r^t - 1)(F/p) < c_2 r^{t(1-\delta)}, \tag{2.30}$$

for a suitably chosen constant c_2. (2.30) yields $F/p < (2c_2)r^{-\delta t}$. Pick $k \geq 3$ to be an *odd* integer large enough so that $(1 - 1/c)/(k-2) < \delta$. We now have

$$F/p < r^{-\frac{(1-1/c)t}{(k-2)}} \tag{2.31}$$

for all sufficiently large values of p. Combining $p > r^{t/c}$ with (2.31) we get

$$\frac{F}{|C_r^p|} < \left(\frac{|C_r^p|}{r^t}\right)^{1/(k-2)},$$

and the application of lemma 2.27 together with an observation that for odd k there always exist $\{\sigma_i\}_{i\in[k]}$ in $\mathbb{F}_r^*$, where $\sum_{i\in[k]} \sigma_i \neq 0$ conclude the proof. ■

2.6.3 Summary

We now summarize our sufficient conditions on p and r that yield algebraic niceness of $\langle r\rangle \subseteq \mathbb{F}_p^*$ over $\mathbb{F}_r$. Combining lemmas 2.24 and 2.25 we get

Lemma 2.30 *Suppose $p = 2^t - 1$ is a Mersenne prime; then $\langle 2\rangle \subseteq \mathbb{F}_p^*$ is three algebraically nice over $\mathbb{F}_2$.*

Using lemma 2.26 instead of lemma 2.25 (in combination with lemma 2.24) we get a weaker sufficient condition.

Lemma 2.31 *Suppose p and r are distinct primes such that $\mathrm{ord}_p(r) \leq (4/3)\log_r p$; then $\langle r\rangle \subseteq \mathbb{F}_p^*$ is three algebraically nice over $\mathbb{F}_r$.*

Finally, combining lemmas 2.24 and 2.29 we get

Lemma 2.32 *For every $c > 0$ and prime r there exists an integer $k = k(c, r)$ such that the following implication holds. If $p \neq r$ is a prime and $\mathrm{ord}_p(r) < c\log_r p$ then $\langle r\rangle \subseteq \mathbb{F}_p^*$ is k algebraically nice over $\mathbb{F}_r$.*

2.7 Results

In what follows we put the results of the previous sections together and summarize our improvements in upper bounds for locally decodable codes.

In section 2.7.1 we present our results for the narrow case of three query binary codes. First we show that given a single Mersenne prime $p = 2^t - 1$ one can design three query binary LDCs of length $\exp(n^{1/t})$ for every message length n. Secondly we review the achievements of the centuries-old study of Mersenne primes, and present vast explicit improvements upon the earlier work.

In section 2.7.2 we present the general form our our results. We show that if r is a prime and $r^t - 1$ has a polynomially large prime factor $p \geq r^{\gamma t}$; then for every message length n there exists a $k(\gamma)$-query r-ary LDC of length $\exp(n^{1/t})$. The query complexity of our codes depends on the size of the largest prime factor of $r^t - 1$, and the length of our codes depends on the size of $r^t - 1$ itself. The larger is the largest prime factor the smaller is the query complexity. The larger is $r^t - 1$ the shorter are the codes.

2.7.1 Results for three query binary codes

Combining proposition 2.20 with lemma 2.22 and lemma 2.30 we conclude that every Mersenne prime $p = 2^t - 1$ yields a family of 3-query LDCs of length $\exp\left(n^{1/t}\right)$:

Theorem 2.33 *Suppose $p = 2^t - 1$ is a Mersenne prime; then for every message length n there exists a binary linear code of length $\exp\left(n^{1/t}\right)$ that is $(3, \delta, 6\delta)$ locally decodable for all δ.*

Mersenne primes have been a popular object of study in number theory for the last few centuries. The largest known Mersenne prime (as of June, 2007) is $p = 2^{32,582,657} - 1$. It was discovered by C. Cooper and S. Boone [CB] on September 4, 2006. Plugging p into theorem 2.33 we get

Theorem 2.34 *For every message length n there exists a binary linear code of length $\exp\left(n^{1/32,582,657}\right)$ that is $(3, \delta, 6\delta)$ locally decodable for all δ.*

It has often been conjectured that the number of Mersenne primes is infinite. If this conjecture holds we get three query locally decodable codes of subexponential length *for infinitely many* message lengths n. To prove this we first combine proposition 2.19 with lemmas 2.21 and 2.30 to obtain

Lemma 2.35 *Let $p = 2^t - 1$ be a Mersenne prime and $m \geq p - 1$ be an integer. Let $m' = \binom{m-1+(p-1)/t}{(p-1)/t}$. There exists a binary linear code encoding $n = \binom{m}{p-1}$ bits to $p^{m'}$ bits that is $(3, \delta, 6\delta)$ locally decodable code for all δ.*

Now we are proceed to constructing a family of three query LDCs of subexponential length.

Theorem 2.36 *Suppose that the number of Mersenne primes is infinite; then for infinitely many values of message length n there exists a binary linear code of length $\exp\left(n^{O\left(\frac{1}{\log\log n}\right)}\right)$ that is $(3, \delta, 6\delta)$ locally decodable for all δ.*

Proof: Given a Mersenne prime p, set $m = 2^p$. Substituting m and p into lemma 2.35 and making some basic manipulations we conclude that there exists a $(3, \delta, 6\delta)$ locally decodable code encoding $n = m^{\Theta(\log m)}$ bits to $N = \exp\left(m^{O\left(\frac{\log m}{\log\log m}\right)}\right)$ bits. An observation that $\log\log n = \Theta(\log\log m)$ completes the proof. ■

Lenstra, Pomerance, and Wagstaff [LPW, Pom80, Wag83] have made the following conjecture regarding the density of Mersenne primes.

Conjecture 2.37 *Let $M(t)$ be the number of Mersenne primes that are less than or equal to $2^t - 1$; then*

$$\lim_{t\to\infty} M(t)/\log_2 t = e^\gamma,$$

where $\gamma \approx 0.577$ is the Euler-Mascheroni constant.

If the conjecture above holds we get three query locally decodable codes of subexponential length *for all* message lengths n.

Theorem 2.38 *Let ϵ be a positive constant. Suppose the conjecture 2.37 holds; then for every message length n there exists a binary linear code of length $\exp\left(n^{O\left(\frac{1}{\log^{1-\epsilon}\log n}\right)}\right)$ that is $(3, \delta, 6\delta)$ locally decodable for all δ.*

Proof: Conjecture 2.37 implies that for all sufficiently large integers z there is a Mersenne prime between $2^{\log^{1-\epsilon} z}$ and z. Assume n is sufficiently large. Pick a Mersenne prime p from the interval $\left[2^{\log^{1-\epsilon}\sqrt{\log n}}, \sqrt{\log n}\right]$. Let m be the smallest integer such that $n \leq \binom{m}{p-1}$. Note that $m = pn^{\Theta(1/p)}$. Given an n-bit message x we pad it with zeros to length $\binom{m}{p-1}$ and use the code from lemma 2.35 to encode x into a codeword of length $p^{m'}$ for $m' = \left(n^{1/p}\log p\right)^{O(p/\log p)}$. It remains to notice that $\log m' = O\left(\frac{\log n}{\log p} + \frac{p\log\log p}{\log p}\right) = O\left(\frac{\log n}{\log^{1-\epsilon}\log n}\right)$. ■

2.7.2 Results for general codes

For an integer m let $P(m)$ denote the largest prime factor of m. Our first theorem gets 3-query r-ary LDCs from numbers $m = r^t - 1$ such that $P(m) > m^{3/4}$.

Theorem 2.39 *Let r be a prime. Suppose $P(r^t - 1) > r^{0.75t}$; then for every message length n there exists a three query r-ary code of length $\exp(n^{1/t})$ that is $(3, \delta, 3\delta r/(r-1))$ locally decodable for all δ.*

Proof: Let $P(r^t - 1) = p$. Observe that $p \mid r^t - 1$ and $p > r^{0.75t}$ yield $\mathrm{ord}_p(r) < (4/3)\log_r p$. Combining lemmas 2.31 and 2.22 with proposition 2.20 we obtain the statement of the theorem. ■

As an example application of theorem 2.39 one can observe that $P(2^{23}-1) = 178481 > 2^{(3/4)*23} \approx 155872$ yields a family of three query locally decodable codes of length $\exp(n^{1/23})$. Theorem 2.39 immediately yields:

Theorem 2.40 *Let r be a prime. Suppose for infinitely many t we have $P(r^t - 1) > r^{0.75t}$; then for every $\epsilon > 0$, for every message length n there exists a three query r-ary code of length $\exp(n^{\epsilon})$ that is $(3, \delta, 3\delta r/(r-1))$ locally decodable for all δ.*

The next theorem gets constant query LDCs from numbers $m = r^t - 1$ with prime factors larger than m^{γ} for every value of γ.

Theorem 2.41 *Let r be a prime. For every $\gamma > 0$ there exists an integer $k = k(\gamma, r)$ such that the following implication holds. Suppose $P(r^t - 1) > r^{\gamma t}$; then for every message length n there exists a k query r-ary code of length $\exp(n^{1/t})$ that is $(k, \delta, \delta k r/(r-1))$ locally decodable for all δ.*

Proof: Let $P(r^t - 1) = p$. Observe that $p \mid r^t - 1$ and $p > r^{\gamma t}$ yield $\mathrm{ord}_p(r) < (1/\gamma)\log_r p$. Combining lemmas 2.32 and 2.22 with proposition 2.20 we obtain the statement of the theorem. ■

As an immediate corollary we get:

Theorem 2.42 *Let r be a prime. Suppose for some $\gamma > 0$ and infinitely many t we have $P(r^t - 1) > r^{\gamma t}$; then there is a fixed k such that for every $\epsilon > 0$, for every message length n there exists a k-query r-ary code of length $\exp(n^\epsilon)$ that are $(k, \delta, \delta kr/(r-1))$ locally decodable for all δ.*

Chapter 3

Limitations of the point removal method

In the previous chapter we gave a detailed exposition of the point removal method for constructing locally decodable codes and obtained vast improvement upon the earlier work. Our most general result (theorem 2.41) said that if r is a prime and $r^t - 1$ has a polynomially large prime factor $p \geq r^{\gamma t}$; then there exists a family of $k(\gamma)$-query r-ary LDC of length $\exp(n^{1/t})$.

In this chapter we prove a partial converse of theorem 2.41. Namely, we show that if for some fixed k and all $\epsilon > 0$ one can use the point removal method to obtain a family of r-ary k-query LDCs of length $\exp(n^{\epsilon})$; then infinitely many numbers of the form $r^t - 1$ prime factors larger than known currently. Our result identifies the problem of establishing strong lower bounds for the size of the largest prime factors of (Mersenne type) numbers $r^t - 1$ as the current barrier for further progress on LDC constructions via the point removal method.

3.1 Attaining subexponential length requires a nice sequence

Point removal method. We start with a high-level review of our construction of

LDCs via point removal that was earlier given in chapter 2. There have been two steps to the construction. First, in propositions 2.19 and 2.20, we argued that every subset S of a finite field $\mathbb{F}_q$ that exhibits two properties (namely, t combinatorial niceness and k algebraic niceness) yields a family of k-query locally decodable codes of length $\exp(n^{1/t})$. Next we came up with a specific set that is simultaneously combinatorially and algebraically nice. Let $P(m)$ denote the largest prime factor of an integer m. In lemmas 2.22 and 2.32, we argued that for a prime r such that $p = P(r^t - 1) \geq r^{\gamma t}$, the multiplicative subgroup of $\mathbb{F}_p^*$ generated by r (that we denote by $\langle r \rangle$) is simultaneously t combinatorially nice and $k(\gamma)$ algebraically nice. A combination of the two steps above led to theorems 2.39 and 2.41, saying that every number of the form $r^t - 1$ that has a polynomially large prime factor gives rise to a family of short locally decodable codes. Instantiating theorem 2.39 with the largest known Mersenne prime we got three query binary codes of length $\exp(n^{1/32,582,657})$ presenting a vast improvement upon the earlier work.

Point removal and bounds for $P(r^t - 1)$. Given the magnitude of our improvements it is natural to ask if the same technique could lead to even shorter codes. Specifically, one could ask whether it is possible to use the same ideas to obtain families of k-query codes of length $\exp(n^{\epsilon})$ for some fixed k and all $\epsilon > 0$. In theorems 2.40 and 2.42 we identified the number theoretic claim that we need to get such codes via the sets $\langle r \rangle \subseteq \mathbb{F}_p^*$. Specifically, we argued that we need a theorem saying that for some $\gamma > 0$ and for some prime r there exist infinitely many t such that $P(2^t - 1) \geq 2^{\gamma t}$. Unfortunately, proving such a strong lower bound on the size of the largest prime factors of Mersenne type numbers is far beyond what number theorists can do nowadays. Lower bounds for $P(r^t - 1)$ have received a considerable amount of attention, especially in the case of $r = 2$ [Sch62, Ste74, ES76, Ste77, MW00, MP04, FLS07]. The strongest result to date is due to Stewart [Ste77]. It says that for all integers t ignoring a set of asymptotic density zero, and for all functions $\epsilon(t) > 0$ where $\epsilon(t)$

tends to zero monotonically and arbitrarily slowly:

$$P(2^t - 1) > \epsilon(t)t\,(\log t)^2\,/\log\log t. \tag{3.1}$$

Although the bound (3.1) may seem extremely weak in light of the conjecture saying that the number of Mersenne primes is infinite there are no better bounds known to hold for infinitely many values of t, unless one is willing to accept some number theoretic conjectures [MW00, MP04].

Our results. In this chapter we show that the need for stronger lower bounds for $P(r^t - 1)$ arises not because of the poor choice of the set $S = \langle r \rangle \subseteq \mathbb{F}_p^*$ but rather is essential to the whole point removal method. Specifically we show that if for some prime r, constant k and every $\epsilon > 0$ one can pick a finite field $\mathbb{F}$ and a set $S \subseteq \mathbb{F}^*$ to (unconditionally) obtain a family of k-query LDCs of length $\exp(n^\epsilon)$ via proposition 2.20; then for infinitely many t we have

$$P(r^t - 1) \geq (t/2)^{1+1/(k-2)}. \tag{3.2}$$

To get a feeling whether the bound above represents a serious barrier to further progress on upper bounds for LDCs via point removal, note that in case $r = 2$ the bound (3.2) is substantially stronger than what is currently known unconditionally (3.1) (for any $k \geq 3$).

We now introduce the notion of a k-nice sequence of subsets of finite fields.

Definition 3.1 *Let r be a prime. We say that a sequence $\left\{S_i \subseteq \mathbb{F}_{q_i}^*\right\}_{i\geq 1}$ of subsets of finite fields is k-nice over $\mathbb{F}_r$ if every S_i is k algebraically nice over $\mathbb{F}_r$ and $t(i)$ combinatorially nice, for some integer-valued monotonically increasing function t.*

It is easy to verify that one needs to exhibit a sequence that is k-nice over $\mathbb{F}_r$ in order to obtain k-query r-ary LDCs of length $\exp(n^\epsilon)$ for some fixed k and every $\epsilon > 0$ via proposition 2.20.

In what follows we show how the existence of a k-nice sequence over $\mathbb{F}_r$ implies that infinitely many numbers $r^t - 1$ have large prime factors. Recall that C_r^p denotes the set of p-th roots of unity in $\overline{\mathbb{F}}_r$. Also recall our notion of a *nontrivial k-dependence:* there is a nontrivial k-dependence in C_r^p if there exist $\{\zeta_i\}_{i\in[k]} \subseteq C_r^p$ and $\{\sigma_i\}_{i\in[k]} \subseteq \mathbb{F}_r$ such that $\sum_i \sigma_i\zeta_i = 0$ and $\sum_i \sigma_i \neq 0$.

Our argument proceeds in two steps. In section 3.2 we show that a k-nice sequence over $\mathbb{F}_r$ yields an infinite sequence of primes $\{p_i\}_{i\geq 1}$, where every $C_r^{p_i}$ contains a nontrivial k-dependence. In sections 3.3 and 3.4 we show that C_r^p contains a nontrivial short dependence only if p is a large factor of a number $r^t - 1$.

3.2 A nice sequence yields short dependencies between p-th roots

Our argument in this section comes in three steps. In subsection 3.2.1 we study algebraically nice subsets of $\mathbb{F}_q^*$. In subsection 3.2.2 we study combinatorially nice subsets of $\mathbb{F}_q^*$. Finally in subsection 3.2.3 we show how an interplay between the structural properties of algebraically and combinatorially nice subsets translates nice sequences over $\mathbb{F}_r$ into infinite families of primes p with short non-trivial dependencies in C_r^p.

3.2.1 Algebraically nice subsets of $\mathbb{F}_q^*$

We start with a review of the definition of algebraic niceness (definition 2.16). A subset $S \subseteq \mathbb{F}_q^*$ is called k algebraically nice over $\mathbb{F}_r$ if there exist maps S_0, S_1 from $\mathbb{F}_q$ to $\mathbb{F}_r$ such that $\mathrm{supp}(S_0) \neq 0$, $\mathrm{supp}(S_1) \leq k$, $\sum_{\lambda\in\mathbb{F}_q} S_1(\lambda) \neq 0$, and for all $\alpha \in \mathbb{F}_q$, $\beta \in S$:

$$\sum_{\lambda\in\mathbb{F}_q} S_0(\alpha + \beta\lambda)S_1(\lambda) \neq 0.$$

The last constraint can be equivalently formulated as: for all $\alpha \in \mathbb{F}_q$ and $\beta \in S$:

$$\sum_{\lambda \in \mathbb{F}_q} S_0(\lambda) S_1((\lambda - \alpha)\beta^{-1}) \neq 0. \tag{3.3}$$

To proceed we need some notation. Consider a a finite field $\mathbb{F}_q = \mathbb{F}_{p^l}$, where p is prime. Fix a basis $e_1, \ldots, e_l$ of $\mathbb{F}_q$ over $\mathbb{F}_p$. In what follows we often write $(\alpha_1, \ldots, \alpha_l) \in \mathbb{F}_p^l$ to denote $\alpha = \sum_{i=1}^{l} \alpha_i e_i \in \mathbb{F}_q$. Let r be a prime. Let R denote the ring $\mathbb{F}_r[x_1, \ldots, x_l]/(x_1^p - 1, \ldots, x_l^p - 1)$. For $\alpha = (\alpha_1, \ldots, \alpha_l) \in \mathbb{F}_q$ we write x^α to denote the monomial $x_1^{\alpha_1} \ldots x_l^{\alpha_l} \in R$. Consider a natural one to one correspondence between maps $S_1 : \mathbb{F}_q \to \mathbb{F}_r$ and polynomials $\phi_{S_1}(x_1, \ldots, x_l) \in R$.

$$\phi_{S_1}(x_1, \ldots, x_l) = \sum_{\lambda \in \mathbb{F}_q} S_1(\lambda) x^\lambda.$$

It is easy to see that for all maps $S_1 : \mathbb{F}_q \to \mathbb{F}_r$ and for all fixed $\alpha \in \mathbb{F}_q$, $\beta \in \mathbb{F}_q^*$:

$$\begin{gathered} \phi_{S_1((\lambda-\alpha)\beta^{-1})}(x_1, \ldots, x_l) = \sum_{\lambda \in \mathbb{F}_q} S_1((\lambda - \alpha)\beta^{-1}) x^\lambda = \\ \sum_{\lambda \in \mathbb{F}_q} S_1(\lambda) x^{\alpha + \beta\lambda} = x_1^{\alpha_1} \ldots x_l^{\alpha_l} \phi_{S_1(\lambda/\beta)}(x_1, \ldots, x_l). \end{gathered} \tag{3.4}$$

Let Γ be a family of maps $\mathbb{F}_q \to \mathbb{F}_r$. It is straightforward to verify that a map $S_0 : \mathbb{F}_q \to \mathbb{F}_r$ satisfies $\sum_{\lambda \in \mathbb{F}_q} S_0(\lambda) S_1(\lambda) = 0$, for every $S_1 \in \Gamma$ if and only if ϕ_{S_0} belongs to $L^\perp$, where L is the linear subspace of R spanned by $\{\phi_{S_1}\}_{S_1 \in \Gamma}$.

Combining the last observation with formulae (3.3) and (3.4) we conclude that a set $S \subseteq \mathbb{F}_q^*$ is k algebraically nice over $\mathbb{F}_r$ if and only if there exists a map $S_1 : \mathbb{F}_q \to \mathbb{F}_r$ such that $\text{supp}(S_1) \leq k$, $\sum\limits_{\lambda \in \mathbb{F}_q} S_1(\lambda) \neq 0$, and the ideal generated by polynomials $\left\{\phi_{S_1(\lambda/\beta)}\right\}_{\{\beta \in S\}}$ is a *proper* ideal of R.

Note that polynomials $\{f_1, \ldots, f_h\}$ generate a proper ideal in R if an only if polynomials $\{f_1, \ldots, f_h, x_1^p - 1, \ldots, x_l^p - 1\}$ generate a proper ideal in $\mathbb{F}_r[x_1, \ldots, x_l]$. Also note that a family of polynomials generates a proper ideal in $\mathbb{F}_r[x_1, \ldots, x_l]$ if and only if it generates a proper ideal in $\overline{\mathbb{F}}_r[x_1, \ldots, x_l]$. Now an application of Hilbert's Nullstellensatz [CLO96, p. 168] implies that a set $S \subseteq \mathbb{F}_q^*$ is k algebraically nice over $\mathbb{F}_r$ if and only if there exists a map $S_1 : \mathbb{F}_q \to \mathbb{F}_r$, with $\mathrm{supp}(S_1) \leq k$ and $\sum\limits_{\lambda\in\mathbb{F}_q} S_1(\lambda) \neq 0$, such that the polynomials $\left\{\phi_{S_1(\lambda/\beta)}\right\}_{\{\beta\in S\}}$ and $\{x_i^p - 1\}_{1\leq i\leq l}$ have a common root in $\overline{\mathbb{F}}_r$.

Lemma 3.2 *Let $\mathbb{F}_q = \mathbb{F}_{p^l}$, where p is prime. Suppose $\mathbb{F}_q$ contains a subset that is nonempty and k algebraically nice over $\mathbb{F}_r$; then there exists a nontrivial k dependence in C_r^p.*

Proof: Assume $S \subseteq \mathbb{F}_q^*$ is nonempty and k algebraically nice over $\mathbb{F}_r$. The discussion above implies that there exists a map $S_1 : \mathbb{F}_q \to \mathbb{F}_r$ such that $\mathrm{supp}(S_1) \leq k$, $\sum_{\lambda\in\mathbb{F}_q} S_1(\lambda) \neq 0$ and all polynomials $\left\{\phi_{S_1(\lambda/\beta)}\right\}_{\{\beta\in S\}}$ vanish at some $(\zeta_1, \ldots, \zeta_l) \in \left(C_r^p\right)^l$. Fix an arbitrary $\beta_0 \in S$, and note that C_r^p is closed under multiplication. Thus,

$$\phi_{S_1(\lambda/\beta_0)}(\zeta_1, \ldots, \zeta_l) = 0 \tag{3.5}$$

yields a nontrivial k dependence in C_r^p. ■

Note that lemma 3.2 does not suffice to prove that a k-nice sequence $\left\{S_i \subseteq \mathbb{F}_{q_i}^*\right\}_{i\geq 1}$ over $\mathbb{F}_r$ yields infinitely many primes p with short nontrivial k dependencies in C_r^p. We need to argue that the set $\{\mathrm{char}\mathbb{F}_{q_i}\}_{i\geq 1}$ can not be finite.

To proceed, we need some more notation. Recall that $q = p^l$ and p is prime. For $x \in \mathbb{F}_q$ let $\mathrm{Tr}(x) = x + \ldots + x^{p^{l-1}} \in \mathbb{F}_p$ denote the trace of x. For $\gamma \in \mathbb{F}_q, c \in \mathbb{F}_p^*$ we call the set $\pi_{\gamma,c} = \{x \in \mathbb{F}_q \mid \mathrm{Tr}(\gamma x) = c\}$ a *proper affine hyperplane* of $\mathbb{F}_q$.

Lemma 3.3 *Let $\mathbb{F}_q = \mathbb{F}_{p^l}$, where p is prime. Suppose $S \subseteq \mathbb{F}_q^*$ is k algebraically nice over $\mathbb{F}_r$; then there exist $h \leq p^k$ proper affine hyperplanes $\{\pi_{\gamma_r,c_r}\}_{1\leq r\leq h}$ of $\mathbb{F}_q$ such that $S \subseteq \bigcup\limits_{r=1}^{h} \pi_{\gamma_r,c_r}$.*

Proof: Discussion preceding lemma 3.2 implies that there exists a map $S_1 : \mathbb{F}_q \to \mathbb{F}_r$ with $\mathrm{supp}(S_1) \leq k$ and $\sum_{\lambda\in\mathbb{F}_q} S_1(\lambda) \neq 0$ such that all polynomials $\left\{\phi_{S_1(\lambda/\beta)}\right\}_{\{\beta\in S\}}$ vanish at some $(\zeta_1, \ldots, \zeta_l) \in \left(C_r^p\right)^l$. Let ζ be a generator of C_p. For every $1 \leq i \leq l$

pick $\omega_i \in \mathbb{Z}_p$ such that $\zeta_i = \zeta^{\omega_i}$. Let $T = \{\tau \in \mathbb{F}_q \mid S_1(\tau) \neq 0\}$. Put $T = \{\tau_1, \ldots, \tau_{k'}\}$. Clearly, $k' \leq k$. For every $\beta \in S$, $\phi_{S_1(\lambda/\beta)}(\zeta_1, \ldots, \zeta_l) = 0$ yields

$$\sum_{\lambda=(\lambda_1,\ldots,\lambda_l)\in\beta T} S_1(\lambda/\beta)\zeta^{\sum_{i=1}^{l}\lambda_i\omega_i} = 0. \tag{3.6}$$

Observe that for fixed values $\{\omega_i\}_{1\leq i\leq l} \in \mathbb{Z}_p$ the map $D(\lambda) = \sum_{i=1}^{l} \lambda_i\omega_i$ is a linear map from $\mathbb{F}_q$ to $\mathbb{F}_p$. It is not hard to prove that every such map can be expressed as $D(\lambda) = \mathrm{Tr}(\delta\lambda)$ for an appropriate choice of $\delta \in \mathbb{F}_q$. Therefore we can rewrite (3.6) as

$$\sum_{\lambda\in\beta T} S_1(\lambda/\beta)\zeta^{\mathrm{Tr}(\delta\lambda)} = \sum_{\tau\in T} S_1(\tau)\zeta^{\mathrm{Tr}(\delta\beta\tau)} = 0. \tag{3.7}$$

Let $W = \left\{(w_1, \ldots, w_{k'}) \in \mathbb{Z}_p^{k'} \mid S_1(\tau_1)\zeta^{w_1} + \ldots + S_1(\tau_{k'})\zeta^{w_{k'}} = 0\right\}$ denote the set of exponents of k'-dependencies between powers of ζ. Clearly, $|W| \leq p^k$. Identity (3.7) implies that every $\beta \in S$ satisfies

$$\begin{cases} \mathrm{Tr}((\delta\tau_1)\beta) & = & w_1, \\ \vdots & & \\ \mathrm{Tr}((\delta\tau_{k'})\beta) & = & w_{k'}; \end{cases} \tag{3.8}$$

for an appropriate choice of $(w_1, \ldots, w_{k'}) \in W$. Note that the all-zeros vector does not lie in W since $\sum_{\tau\in T} S_1(\tau) = \sum_{\lambda\in\mathbb{F}_q} S_1(\lambda) \neq 0$. Therefore at least one of the identities in (3.8) has a non-zero right-hand side, and defines a proper affine hyperplane of $\mathbb{F}_q$. Collecting one such hyperplane for every element of W we get a family of $|W|$ proper affine hyperplanes containing every element of S. ■

3.2.2 Combinatorially nice subsets of $\mathbb{F}_q^*$

Lemma 3.3 gives us some insight into the structure of algebraically nice subsets of $\mathbb{F}_q$. Our next goal is to develop an insight into the structure of combinatorially nice subsets. We start by reviewing some relations between tensor and dot products of vectors. For vectors $u \in \mathbb{F}_q^m$ and $v \in \mathbb{F}_q^n$ let $u \otimes v \in \mathbb{F}_q^{mn}$ denote the tensor product of

u and v. Coordinates of $u \otimes v$ are labelled by all possible elements of $[m] \times [n]$ and $(u \otimes v)_{i,j} = u_i v_j$. Also, let $u^{\otimes l}$ denote the l-the tensor power of u and $u \circ v$ denote the concatenation of u and v. The following identity is standard. For any $u, x \in \mathbb{F}_q^m$ and $v, y \in \mathbb{F}_q^n$:

$$(u \otimes v, x \otimes y) = \sum_{i \in [m], j \in [n]} u_i v_j x_i y_j = \left(\sum_{i \in [m]} u_i x_i \right) \left(\sum_{j \in [n]} v_j y_j \right) = (u, x)(v, y). \quad (3.9)$$

In what follows we need a generalization of identity (3.9). Let $f(x_1, \ldots, x_h) = \sum_i c_i x_1^{\alpha_1^i} \ldots x_h^{\alpha_h^i}$ be a polynomial in $\mathbb{F}_q[x_1, \ldots, x_h]$. Given f we define $\bar{f} \in \mathbb{F}_q[x_1, \ldots, x_h]$ by $\bar{f} = \sum_i x_1^{\alpha_1^i} \ldots x_h^{\alpha_h^i}$, i.e., we simply set all nonzero coefficients of f to 1. For vectors $u_1, \ldots, u_h$ in $\mathbb{F}_q^m$ define

$$f(u_1, \ldots, u_h) = \circ_i \; c_i u_1^{\otimes \alpha_1^i} \otimes \ldots \otimes u_h^{\otimes \alpha_h^i}. \quad (3.10)$$

Note that to obtain $f(u_1, \ldots, u_h)$ we replaced products in f by tensor products and addition by concatenation. Clearly, $f(u_1, \ldots, u_h)$ is a vector whose length may be larger than m.

Claim 3.4 *For every $f \in \mathbb{F}_q[x_1, \ldots, x_h]$ and $u_1, \ldots, u_h, v_1, \ldots, v_h \in \mathbb{F}_q^m$:*

$$\left(f(u_1, \ldots, u_h), \bar{f}(v_1, \ldots, v_h) \right) = f((u_1, v_1), \ldots, (u_h, v_h)). \quad (3.11)$$

Proof: Let $\mathbf{u} = (u_1, \ldots, u_h)$ and $\mathbf{v} = (v_1, \ldots, v_h)$. Observe that if (3.11) holds for polynomials f_1 and f_2 defined over disjoint sets of monomials then it also holds for $f = f_1 + f_2$:

$$\left(f(\mathbf{u}), \bar{f}(\mathbf{v}) \right) = \left((f_1 + f_2)(\mathbf{u}), (\bar{f_1} + \bar{f_2})(\mathbf{v}) \right) = \left(f_1(\mathbf{u}) \circ f_2(\mathbf{u}), \bar{f_1}(\mathbf{v}) \circ \bar{f_2}(\mathbf{v}) \right) =$$

$$f_1 \left((u_1, v_1), \ldots, (u_h, v_h) \right) + f_2 \left((u_1, v_1), \ldots, (u_h, v_h) \right) = f \left((u_1, v_1), \ldots, (u_h, v_h) \right).$$

Therefore it suffices to prove (3.11) for monomials $f = c x_1^{\alpha_1} \ldots x_h^{\alpha_h}$. It remains to notice identity (3.11) for monomials $f = c x_1^{\alpha_1} \ldots x_h^{\alpha_h}$ follows immediately from formula (3.9) using induction on $\sum_{i=1}^h \alpha_i$. ■

The next lemma bounds combinatorial niceness of certain subsets of $\mathbb{F}_q^*$.

Lemma 3.5 *Let $\mathbb{F}_q = \mathbb{F}_{p^l}$, where p is prime. Let $S \subseteq \mathbb{F}_q^*$. Suppose there exist h proper affine hyperplanes $\{\pi_{\gamma_r, c_r}\}_{1 \leq r \leq h}$ of $\mathbb{F}_q$ such that $S \subseteq \bigcup_{r=1}^{h} \pi_{\gamma_r, c_r}$; then S is at most $h(p-1)$ combinatorially nice.*

Proof: Assume S is t combinatorially nice. This implies that for some $c > 0$ and every m there exist two $n = \lfloor cm^t \rfloor$-sized collections of vectors $\{u_i\}_{i\in[n]}$ and $\{v_i\}_{i\in[n]}$ in $\mathbb{F}_q^m$, such that:

- For all $i \in [n]$, $(u_i, v_i) = 0$;
- For all $i, j \in [n]$ such that $i \neq j$, $(u_j, v_i) \in S$.

For a vector $u \in \mathbb{F}_q^m$ and integer e let u^e denote a vector resulting from raising every coordinate of u to the power e. For every $i \in [n]$ and $r \in [h]$ define vectors $u_i^{(r)}$ and $v_i^{(r)}$ in $\mathbb{F}_q^{ml}$ by

$$u_i^{(r)} = (\gamma_r u_i) \circ (\gamma_r u_i)^p \circ \ldots \circ (\gamma_r u_i)^{p^{l-1}} \quad \text{and} \quad v_i^{(r)} = v_i \circ v_i^p \circ \ldots \circ v_i^{p^{l-1}}. \tag{3.12}$$

Note that for every $r_1, r_2 \in [h]$, $v_i^{(r_1)} = v_i^{(r_2)}$. It is straightforward to verify that for every $i, j \in [n]$ and $r \in [h]$:

$$\left(u_j^{(r)}, v_i^{(r)}\right) = \mathrm{Tr}(\gamma_r(u_j, v_i)). \tag{3.13}$$

Combining (3.13) with the fact that S is covered by proper affine hyperplanes $\{\pi_{\gamma_r, c_r}\}_{r\in[h]}$ we conclude that

- For all $i \in [n]$ and $r \in [h]$, $\left(u_i^{(r)}, v_i^{(r)}\right) = 0$;
- For all $i, j \in [n]$ such that $i \neq j$, there exists $r \in [h]$ such that $\left(u_j^{(r)}, v_i^{(r)}\right) \in \mathbb{F}_p^*$.

Pick $g(x_1, \ldots, x_h) \in \mathbb{F}_p[x_1, \ldots, x_h]$ to be a homogeneous degree h polynomial such that for $\mathbf{a} = (a_1, \ldots, a_h) \in \mathbb{F}_p^h : g(\mathbf{a}) = 0$ if and only if $\mathbf{a}$ is the all-zeros vector. The existence of such a polynomial g follows from [LN83, Example 6.7]. Set $f = g^{p-1}$.

Note that for $\mathbf{a} \in \mathbb{F}_p^h$: $f(a) = 0$ if $\mathbf{a}$ is the all-zeros vector, and $f(a) = 1$ otherwise. For all $i \in [n]$ define

$$u_i' = f\left(u_i^{(1)}, \ldots, u_i^{(h)}\right) \circ (1) \quad \text{and} \quad v_i' = \bar{f}\left(v_i^{(1)}, \ldots, v_i^{(h)}\right) \circ (-1). \tag{3.14}$$

Note that f and $\bar{f}$ are homogeneous degree $(p-1)h$ polynomials in h variables. Therefore (3.10) implies that for all i vectors u_i' and v_i' have length $m' \leq h^{(p-1)h}(ml)^{(p-1)h}+1$. Combining identities (3.14) and (3.11) and using the properties of dot products between vectors $\left\{u_i^{(r)}\right\}$ and $\left\{v_i^{(r)}\right\}$ discussed above we conclude that for every m there exist two $n = \lfloor cm^t \rfloor$-sized collections of vectors $\{u_i'\}_{i\in[n]}$ and $\{v_i'\}_{i\in[n]}$ in $\mathbb{F}_q^{m'}$, such that:

- For all $i \in [n]$, $(u_i', v_i') = -1$;
- For all $i, j \in [n]$ such that $i \neq j$, $(u_j, v_i) = 0$.

Notice that a family of vectors with such properties exists only if $n \leq m'$, i.e.,

$$\lfloor cm^t \rfloor \leq h^{(p-1)h}(ml)^{(p-1)h} + 1.$$

Given that we can pick m to be arbitrarily large, this implies that $t \leq (p-1)h$. ■

3.2.3 Summary

The next lemma presents the main result of this section.

Lemma 3.6 *Let r be a prime. Suppose there exists a k-nice sequence over $\mathbb{F}_r$; then for infinitely many primes p there exists a nontrivial k dependence in C_r^p.*

Proof: Assume $\left\{S_i \subseteq \mathbb{F}_{q_i}^*\right\}_{i\geq 1}$ is a k-nice sequence over $\mathbb{F}_r$. Let p be a fixed prime. Combining lemmas 3.3 and 3.5 we conclude that every subset $S \subseteq \mathbb{F}_{p^l}^*$ that is k algebraically nice over $\mathbb{F}_r$ is at most $(p-1)p^k$ combinatorially nice. Note that our bound on combinatorial niceness is independent of l. Therefore there are only finitely many extensions of the field $\mathbb{F}_p$ in the sequence $\{\mathbb{F}_{q_i}\}_{i\geq 1}$, and the set $\mathbb{P} = \{\text{char}\mathbb{F}_{q_i}\}_{i\geq 1}$

is infinite. It remains to notice that according to lemma 3.2 for every $p \in \mathbb{P}$ there exists a nontrivial k dependence in C_r^p. ■

In what follows we present necessary conditions for the existence of nontrivial k dependencies in C_r^p. We treat the $k = 3$ case separately since in that case we can use a specialized argument to derive a slightly stronger conclusion.

3.3 k-dependencies between p-th roots: a necessary condition

Lemma 3.7 *Let p and r be primes. Suppose there exists a nontrivial k dependence in C_r^p; then*

$$\mathrm{ord}_p(r) \leq 2p^{1-1/(k-1)}. \tag{3.15}$$

Proof: Let $\{\zeta_i\}_{i\in[k]} \subseteq C_r^p$ and $\{\sigma_i\}_{i\in[k]} \subseteq \mathbb{F}_r$ be such that $\sum_{i\in[k]} \sigma_i\zeta_i = 0$ and $\sum_{i\in[k]} \sigma_i \neq 0$. Let $t = \mathrm{ord}_p(r)$. Note that $C_r^p \subseteq \mathbb{F}_{r^t}$. Note also that all elements of C_r^p other than the multiplicative identity are proper elements of $\mathbb{F}_{r^t}$. Therefore for every $\zeta \in C_r^p$ where $\zeta \neq 1$ and every nonzero $f(x) \in \mathbb{F}_r[x]$ such that $\deg f \leq t-1$ we have: $f(\zeta) \neq 0$.

By multiplying $\sum_{i=1}^k \sigma_i\zeta_i = 0$ through by ζ_k^{-1}, we may reduce to the case $\zeta_k = 1$. Let ζ be the generator of C_r^p. For every $i \in [k-1]$ pick $w_i \in \mathbb{Z}_p$ such that $\zeta_i = \zeta^{w_i}$. We now have $\sum_{i=1}^{k-1} \sigma_i\zeta^{w_i} + \sigma_k = 0$. Set $h = \lfloor (t-1)/2 \rfloor$. Consider the $(k-1)$-tuples:

$$(mw_1 + i_1, \ldots, mw_{k-1} + i_{k-1}) \in \mathbb{Z}_p^{k-1}, \quad \text{for } m \in \mathbb{Z}_p \text{ and } i_1, \ldots, i_{k-1} \in [0, h]. \tag{3.16}$$

Suppose two of these coincide, say

$$(mw_1 + i_1, \ldots, mw_{k-1} + i_{k-1}) = (m'w_1 + i'_1, \ldots, m'w_{k-1} + i'_{k-1}),$$

with $(m, i_1, \ldots, i_{k-1}) \neq (m', i'_1, \ldots, i'_{k-1})$. Set $n = m - m'$ and $j_l = i'_l - i_l$ for $l \in [k-1]$. We now have

$$(nw_1, \ldots, nw_{k-1}) = (j_1, \ldots, j_{k-1})$$

with $-h \leq j_1, \ldots, j_{k-1} \leq h$. Observe that $n \neq 0$, and thus it has a multiplicative inverse $g \in \mathbb{Z}_p$. Consider a polynomial

$$P(z) = \sigma_1 z^{j_1+h} + \ldots + \sigma_{k-1} z^{j_{k-1}+h} + \sigma_k z^h \in \mathbb{F}_r[z].$$

Note that $\deg P \leq 2h \leq t-1$. Note also that $P(1) \neq 0$ and $P(\zeta^g) = 0$. The latter identity contradicts the fact that ζ^g is a proper element of $\mathbb{F}_{r^t}$. This contradiction implies that all $(k-1)$-tuples in (3.16) are distinct. This yields

$$p^{k-1} \geq p \left(\frac{t}{2}\right)^{k-1},$$

which is equivalent to (3.15). ■

3.4 3-dependencies between p-th roots: a necessary condition

In this section we slightly strengthen lemma 3.7 in the special case when $k = 3$. Our argument is loosely inspired by the Agrawal-Kayal-Saxena deterministic primality test [AKS04].

Lemma 3.8 *Let p and r be primes. Suppose there exists a nontrivial three dependence in C_r^p; then*

$$\mathrm{ord}_p(2) \leq ((4/3)p)^{1/2}. \tag{3.17}$$

Proof: Let $\{\zeta_i\}_{i\in[3]} \subseteq C_r^p$ and $\{\sigma_i\}_{i\in[3]} \subseteq \mathbb{F}_r$ be such that $\sum_{i\in[3]} \sigma_i \zeta_i = 0$ and $\sum_{i\in[3]} \sigma_i \neq 0$. Let $t = \mathrm{ord}_p(r)$. Note that $C_r^p \subseteq \mathbb{F}_{r^t}$. Note also that all elements of C_r^p other than the multiplicative identity are proper elements of $\mathbb{F}_{r^t}$. Therefore for every $\zeta \in C_r^p$ where $\zeta \neq 1$ and every nonzero $f(x) \in \mathbb{F}_r[x]$ such that $\deg f \leq t-1$ we have: $f(\zeta) \neq 0$.

Without loss of generality assume $\sigma_1 \neq 0, \sigma_3 = -1$, and $\zeta_3 = 1$. Observe that $\sigma_1\zeta_1 + \sigma_2\zeta_2 = 1$ implies $\left(\sigma_1\zeta_1\zeta_2^{-1} + \sigma_2\right)^p = 1$. Put $\zeta = \zeta_1\zeta_2^{-1}$. Note that $\zeta \neq 1$ (since

$\sum_{i\in[3]} \sigma_i \neq 0$) and $\zeta, \sigma_1\zeta + \sigma_2 \in C_r^p$. Consider the products $\pi_{i,j} = \zeta^i(\sigma_1\zeta + \sigma_2)^j \in C_r^p$ for $0 \leq i, j \leq t-1$. Note that $\pi_{i,j}, \pi_{k,l}$ cannot be the same if $i \geq k$ and $l \geq j$, as then

$$\zeta^{i-k} - (\sigma_1\zeta + \sigma_2)^{l-j} = 0,$$

but the left side has degree less than t. In other words, if $\pi_{i,j} = \pi_{k,l}$ and $(i,j) \neq (k,l)$, then the pairs (i,j) and (k,l) are comparable under termwise comparison. In particular, either $(k,l) = (i+a, j+b)$ or $(i,j) = (k+a, l+b)$ for some pair (a,b) with $\pi_{a,b} = 1$.

We next check that there cannot be two distinct nonzero pairs $(a,b), (a',b')$ with $\pi_{a,b} = \pi_{a',b'} = 1$. As above, these pairs must be comparable; we may assume without loss of generality that $a \leq a', b \leq b'$. The equations $\pi_{a,b} = 1$ and $\pi_{a'-a,b'-b} = 1$ force $a + b \geq t$ and $(a'-a) + (b'-b) \geq t$, so $a' + b' \geq 2t$. But $a', b' \leq t-1$, contradiction.

If there is no nonzero pair (a,b) with $0 \leq a, b \leq t-1$ and $\pi_{a,b} = 1$, then all $\pi_{i,j}$ are distinct, so $p \geq t^2$. Otherwise, as above, the pair (a,b) is unique, and the pairs (i,j) with $0 \leq i,j \leq t-1$ and $(i,j) \not\geq (a,b)$ are pairwise distinct. The number of pairs excluded by the condition $(i,j) \not\geq (a,b)$ is $(t-a)(t-b)$; since $a+b \geq t$, $(t-a)(t-b) \leq t^2/4$. Hence $p \geq t^2 - t^2/4 = 3t^2/4$ as desired. ■

3.5 Summary

In section 3.1 we argued that in order to use the point removal method to obtain k-query locally decodable codes of length $\exp(n^\epsilon)$ over $\mathbb{F}_r$ for some fixed k and all $\epsilon > 0$, one needs to exhibit a sequence of subsets of finite fields that is k-nice over $\mathbb{F}_r$. In what follows we use technical results of the previous sections to show that the existence of a k-nice sequence over $\mathbb{F}_r$ implies that infinitely many Mersenne type numbers $r^t - 1$ have large prime factors.

Theorem 3.9 *Let r be a prime. Suppose there exists a sequence of subsets of finite* *fields that is k-nice over $\mathbb{F}_r$; then for infinitely many values of t we have*

$$P(r^t - 1) \geq (t/2)^{1+1/(k-2)}. \tag{3.18}$$

Proof: Using lemmas 3.6 and 3.7 we conclude that a k-nice sequence yields infinitely many primes p such that $\mathrm{ord}_p(r) \leq 2p^{1-1/(k-1)}$. Let p be such a prime and $t = \mathrm{ord}_p(r)$. Then $P(r^t - 1) \geq (t/2)^{1+1/(k-2)}$. ■

A combination of lemmas 3.6 and 3.8 yields a slightly stronger bound for the special case of 3-nice sequences.

Theorem 3.10 *Let r be a prime. Suppose there exists a sequence of subsets of finite fields that is three nice over $\mathbb{F}_r$; then for infinitely many values of t we have*

$$P(2^t - 1) \geq (3/4)t^2. \tag{3.19}$$

We would like to remind the reader that although (in case $r = 2$) the lower bounds for $P(r^t - 1)$ given by (3.18) and (3.19) are extremely weak light of the widely accepted conjecture saying that the number of Mersenne primes is infinite, they are substantially stronger than what is currently known unconditionally (3.1).

3.6 Conclusions

Our result in this chapter shows that any attempts to obtain locally decodable codes of length $\exp(n^\epsilon)$ for some fixed query complexity and all $\epsilon > 0$ via the point removal method (i.e., via proposition 2.20) require progress on an old number theory problem. Therefore obtaining such codes using this technique seems unlikely in the near future.

Our result can be used to direct the efforts of researchers looking for better constructions of locally decodable codes. Specifically, the author hopes [Yek06, section 7] that although point removal in finite fields has reached a solid barrier, there still may be room for further progress via generalizations of the point removal idea to suitably chosen finite commutative rings.

Chapter 4

Private information retrieval

In this chapter we study the communication complexity of private information retrieval schemes and obtain both upper and lower bounds. The chapter consists of two uneven parts.

The first part of the chapter (section 4.2) deals with upper bounds. We use the point removal method from chapter 2 to obtain a new generation of PIR schemes. Our constructions yield tremendous improvements for communication complexity of schemes involving three or more servers.

Sections 4.3–4.5 constitute the second (lower bounds) part of this chapter. In section 4.3 we introduce a new restricted (bilinear group-based) model of two server PIR. Our model is fairly broad and captures all currently known two server schemes. In section 4.4 we obtain a tight lower bound for communication complexity of bilinear group-based PIR schemes. Finally, in section 4.5 we discuss possible interpretations of our lower bound.

4.1 Preliminaries

We model the database by an n-long r-ary string. A k-server PIR scheme involves k servers $\mathcal{S}_1, \ldots, \mathcal{S}_k$ each holding the same the database $x \in [r]^n$, and user $\mathcal{U}$ who knows n and wants to retrieve some value x_i, $i \in [n]$, without revealing the value of

i. We restrict our attention to one round information-theoretic PIR protocols. Below is a formal definition of a PIR scheme.

Definition 4.1 *A k-server PIR protocol is a triplet of non-uniform algorithms $\mathcal{P} = (\mathcal{Q}, \mathcal{A}, \mathcal{C})$. We assume that each algorithm is given n as an advice. At the beginning of the protocol, the user $\mathcal{U}$ tosses random coins and obtains a random string rand. Next $\mathcal{U}$ invokes $\mathcal{Q}(i, rand)$ to generate a k-tuple of queries $(que_1, \ldots, que_k)$. For $j \in [k]$, $\mathcal{U}$ sends que_j to $\mathcal{S}_j$. Each server $\mathcal{S}_j$, $j \in [k]$ responds with an answer $ans_j = \mathcal{A}(j, x, que_j)$. (We assume without loss of generality that servers are deterministic.) Finally, $\mathcal{U}$ computes its output by applying the reconstruction algorithm $\mathcal{C}(ans_1, \ldots, ans_k, i, rand)$. A protocol as above should satisfy the following requirements:*

- **Correctness :** *For any n, $x \in [r]^n$ and $i \in [n]$, $\mathcal{U}$ outputs the correct value of x_i with probability 1 (where the probability is over the random strings rand).*

- **Privacy :** *Each server individually learns no information about i. More precisely, we require that for any n and for any $j \in [k]$, the distributions $que_j(i, rand)$ are identical for all values $i \in [n]$.*

The *communication complexity* of a PIR protocol $\mathcal{P}$, is a function of n measuring the total number of bits communicated between the user and the servers, maximized over all choices of $x \in [r]^n$, $i \in [n]$, and random inputs.

In the special case when r is a prime power and the elements of the alphabet $[r]$ are in one to one correspondence with the elements of the finite field $\mathbb{F}_r$ it makes sense to talk about *linear* private information retrieval schemes [GKST02]. A linear PIR scheme is a PIR scheme, where the answer function $\mathcal{A}(j, x, que_j)$ is linear in x over $\mathbb{F}_r$ for arbitrary fixed values of j and que_j. In other words, every coordinate of an answer is a certain linear combination of the database values.

4.2 From LDCs to PIRs

In this section we present our improvements of upper bounds for communication complexity of private information retrieval.

Our improvements follow via a (relatively simple) reduction that turns the constructions of locally decodable codes presented in chapter 2 into constructions of PIR schemes. Note that there are known generic procedures [KT00] to convert LDCs into PIRs. However a simple application of such a procedure to our LDCs would either yield a PIR protocol with perfect privacy, but small probability of error, or a PIR protocol with perfect correctness and some slight privacy leakage. Fortunately, it is possible to achieve both perfect privacy and perfect correctness simultaneously via a specially designed argument.

We now turn to lemma 2.17 that is the core lemma of chapter 2 translating nice subsets of finite fields to codes, and show how a minor modification to the proof of that lemma allows us to build PIR schemes from nice sets. Before we proceed we slightly strengthen the definition of combinatorially nice sets.

Definition 4.2 *Let q be a prime power. A set $S \subseteq \mathbb{F}_q^*$ is called (m,n) normally combinatorially nice if there exist two families of vectors $\{u_1, \ldots, u_n\}$, $\{v_1, \ldots, v_n\}$ and a vector e in $\mathbb{F}_q^m$, such that*

- *For all $i \in [n]$, $(u_i, v_i) = 0$;*
- *For all $i, j \in [n]$ such that $i \neq j$, $(u_j, v_i) \in S$;*
- *For all $i \in [n]$, $(u_i, e) \neq 0$.*

Clearly, every normally combinatorially nice set is combinatorially nice. The converse also holds for all specific combinatorially nice sets that have been considered in this thesis. Our only construction of combinatorially nice sets is given by lemma 2.21. It is straightforward to verify that the all-ones vector $e \in \mathbb{F}_p^m$ is non-orthogonal to every vector u_i considered in that lemma. Our proof of lemma 4.3 assumes reader's familiarity with the proof of lemma 2.17. We write $\log z$ to denote the logarithm base 2.

Lemma 4.3 *Let q be a prime power and r be a prime. Assume $S \subseteq \mathbb{F}_q^*$ is simultaneously (m,n) normally combinatorially nice, and k algebraically nice over $\mathbb{F}_r$. The* *set S yields a one round $\mathbb{F}_r$-linear k-server PIR scheme with questions of bit length $m \log q$ and answers of bit length $q \log r$ that allows private retrieval of coordinate values from an r-ray database of length n.*

Proof: In the preprocessing stage the servers encode the database x with a k query locally decodable code C from lemma 2.17. We are going to use the notation from that lemma. Recall that the coordinates of $C(x)$ are in one to one correspondence with points in $\mathbb{F}_q^m$. In order to decode x_i the user has to query $\mathrm{supp}(S_1)$ locations $\{w + \lambda v_i \mid \lambda : S_1(\lambda) \neq 0\}$ for some $w \in T_i$, where T_i is the union of certain cosets of the hyperplane $\left\{y \in \mathbb{F}_q^m \mid (u_i, y) = 0\right\}$. Unlike the LDC setup in the PIR setup the user can not pick $w \in T_i$ uniformly at random and then query locations $\{w + \lambda v_i \mid \lambda : S_1(\lambda) \neq 0\}$ from $\mathrm{supp}(S_1)$ different servers, since in such case the servers would observe the uniform distribution on T_i rather than the uniform distribution on $\mathbb{F}_q^m$. Here is our way to go around this problem.

Let $e \in \mathbb{F}_q^m$ be such that $(u_i, e) \neq 0$ for all $i \in [n]$. Thus for every $i \in [n]$ and every $w \in \mathbb{F}_q^m$ there is some $\gamma_0 \in \mathbb{F}_q$ such that $w + \gamma_0 e \in T_i$. The user picks $w \in \mathbb{F}_q^m$ uniformly at random and (simultaneously) asks q $\mathrm{supp}(S_1)$-tuples of queries of the from $\{w + \gamma e + \lambda v_i \mid \lambda : S_1(\lambda) \neq 0\}$ for all $\gamma \in \mathbb{F}_q$. For every $\mathrm{supp}(S_1)$-tuple and for every $j \in [\mathrm{Supp}(S_1)]$ the query number j goes to server $\mathcal{S}_j$. (Note that in order to ask all those queries the user needs to communicate only a single point in $\mathbb{F}_q^m$ to each of the servers.) It is easy to verify that in such case each server individually observes a uniform distribution independent of i, while the user always successfully reconstructs x_i from one of the $\mathrm{supp}(S_1)$-tuples of queries. ■

Recall that we have two definitions of combinatorial niceness. One involving two parameters (definition 2.14) and the other involving a single parameter (definition 2.15). Similarly, we would like to have two definitions of normal combinatorial niceness. Therefore we define a set $S \subseteq \mathbb{F}_q^*$ to be t normally combinatorially nice, if there exists a constant $c > 0$ such that for all positive integers m, S is $(m, \lfloor cm^t \rfloor)$ normally combinatorially nice.

Below is a variant of lemma 4.3 involving a single parameter definition of normal combinatorial niceness. The proof is essentially identical to the proof of proposition 2.20 and is omitted.

Lemma 4.4 *Let q be a prime power and r be a prime. Assume $S \subseteq \mathbb{F}_q^*$ is simultaneously t normally combinatorially nice, and k algebraically nice over $\mathbb{F}_r$. For every database length n, the set S yields a one round $\mathbb{F}_r$-linear k-server PIR scheme with questions of bit length $O(n^{1/t})$ and answers of bit length $O(1)$.*

In what follows we use our constructions of normally combinatorially nice and algebraically nice sets from sections 2.5 and 2.6 together with lemmas 4.3 and 4.4 to construct efficient PIR schemes.

In section 4.2.1 we present our results for the narrow case of three server binary PIR schemes. First we show that given a single Mersenne prime $p = 2^t - 1$ one can design three server binary PIRs with $O(n^{1/(t+1)})$ communication to access an n-bit database, for every n. Secondly we use the achievements of the centuries-old study of Mersenne primes, and present vast explicit improvements upon the earlier work.

In section 4.2.2 we present the general form our our results. We show that if r is a prime and $r^t - 1$ has a polynomially large prime factor $p \geq r^{\gamma t}$; then for every database length n there exists a $k(\gamma)$-server r-ary PIR scheme with $O(n^{1/(t+1)})$ communication. The number of servers in our schemes depends on the size of the largest prime factor of $r^t - 1$, and the communication complexity depends on the size of $r^t - 1$ itself. The larger is the largest prime factor the smaller is the number of servers. The larger is $r^t - 1$ the smaller is the communication complexity.

4.2.1 Improved upper bounds for 3-server binary PIRs

Combining lemma 4.4 with lemmas 2.22 and 2.30 we conclude that every Mersenne prime $p = 2^t - 1$ yields a family of 3-server PIRs with $O(n^{1/t})$ communication.

Theorem 4.5 *Let $p = 2^t - 1$ be a fixed Mersenne prime. For every database length n there exists a three server binary PIR protocol with questions of length $O\left(n^{1/t}\right)$ and answers of length $O(1)$.*

A generic balancing technique of [CGKS98, section 4.3] allows to convert any PIR protocol with $O(n^{1/t})$ long questions and $O(1)$ long answers into a new PIR protocol with $O(n^{1/(t+1)})$ total communication. Such a conversion yields

Theorem 4.6 *Let $p = 2^t - 1$ be a fixed Mersenne prime. For every database length n there exists a three server binary PIR protocol with $O\left(n^{1/(t+1)}\right)$ communication.*

Plugging the value of the largest known Mersenne prime $p = 2^{32,582,657} - 1$ into theorem 4.6, we conclude

Theorem 4.7 *For every database length n there exists a three server binary PIR protocol with communication complexity of $O\left(n^{1/32,582,658}\right)$.*

The next two theorems capture the asymptotic parameters of our PIR schemes under the number-theoretic assumptions. Both theorems follow easily by a combination of lemma 4.3 with lemmas 2.22 and 2.30 using the arguments that are essentially identical to the proofs of theorems 2.36 and 2.38.

Theorem 4.8 *Suppose that the number of Mersenne primes is infinite; then for infinitely many database lengths n there exists a three server binary PIR protocol with communication complexity of $n^{O\left(\frac{1}{\log\log n}\right)}$.*

Theorem 4.9 *Let ϵ be a positive constant. Suppose the conjecture 2.37 regarding the density of Mersenne primes holds; then for every database length n there exists a three server binary PIR protocol with communication complexity of $n^{O\left(\frac{1}{\log^{1-\epsilon}\log n}\right)}$.*

4.2.2 Improved upper bounds for general PIRs

For an integer m let $P(m)$ denote the largest prime factor of m. Our first theorem gets 3-server r-ary PIRs from numbers $m = r^t - 1$ with prime factors larger than $m^{3/4}$. The proof is essentially identical to the proof of theorem 2.39 and comes by combining lemmas 4.4, 2.22 and 2.31.

Theorem 4.10 *Let r be a prime. Suppose $P(r^t - 1) > r^{0.75t}$; then for every database length n there exists a three server r-ary PIR protocol with questions of length $O\left(n^{1/t}\right)$ and answers of length $O(1)$.*

Theorem 4.10 immediately yields:

Theorem 4.11 *Let r be a prime. Suppose for infinitely many t we have $P(r^t - 1) > r^{0.75t}$; then for every $\epsilon > 0$ there exists a family of three server r-ary PIR protocols with questions of length $O\left(n^{1/t}\right)$ and answers of length $O(1)$.*

The next theorem gets PIR schemes involving a constant number of servers from numbers $m = r^t - 1$ with prime factors larger than m^γ for every value of γ. The proof is essentially identical to the proof of theorem 2.41 and comes by combining lemmas 4.4, 2.22 and 2.32.

Theorem 4.12 *Let r be a prime. For every $\gamma > 0$ there exists an integer $k = k(\gamma, r)$ such that the following implication holds. Suppose $P(r^t - 1) > r^{\gamma t}$; then for every database length n there exists a k server r-ary PIR protocol with questions of length $O\left(n^{1/t}\right)$ and answers of length $O(1)$.*

As an immediate corollary we get:

Theorem 4.13 *Let r be a prime. Suppose for some $\gamma > 0$ and infinitely many t we have $P(r^t - 1) > r^{\gamma t}$; then there is a fixed k such that for every $\epsilon > 0$ there exists a family of k server r-ary PIR protocols with questions of length $O\left(n^{1/t}\right)$ and answers of length $O(1)$.*

4.3 A combinatorial view of two server PIR

This section begins the second (lower bounds) part of chapter 4. We introduce a new combinatorial interpretation of two server PIR, and identify the models of bilinear PIR and bilinear group based PIR.

We start with some definitions.

Definition 4.14 *A generalized Latin square $Q = GLS[n, T]$ is a square matrix of size T by T over the alphabet $[n] \cup \{*\}$, such that:*

- *For every $i \in [n]$ and $j \in [T]$, there exists a unique $k \in [T]$ such that $Q_{jk} = i$;*

- *For every $i \in [n]$ and $j \in [T]$, there exists a unique $k \in [T]$ such that $Q_{kj} = i$.*

In particular, every row (or column) of a GLS$[n, T]$ contains precisely $(T - n)$ stars. We call the ratio n/T the *density* of a generalized Latin square. It is easy to see that generalized Latin squares of density 1 are simply Latin squares.

Let $Q = \mathrm{GLS}[n, T]$, and let $\sigma : [n] \to [r]$ be an arbitrary map. By Q_σ we denote a matrix of size T by T over the alphabet $[r] \cup \{*\}$, which is obtained from Q by replacing every non-star entry i in Q by $\sigma(i)$. We say that a matrix $C \in [r]^{T \times T}$ is a *completion* of Q_σ if $C_{ij} = (Q_\sigma)_{ij}$ whenever $(Q_\sigma)_{ij} \in [r]$.

For matrices $C \in [r]^{c \times c}$ and $A \in [r]^{l \times l}$ we say that C *reduces* to A if there exist two maps $\pi_1 : [c] \to [l]$ and $\pi_2 : [c] \to [l]$ such that for any $j, k \in [c] : C_{jk} = A_{\pi_1(j), \pi_2(k)}$. Note that we do not impose any restrictions on maps π_1 and π_2, in particular c can be larger then l.

Definition 4.15 *Let $Q = GLS[n, T]$, and $A \in [r]^{l \times l}$. We say that A covers Q, (notation $Q \hookrightarrow A$) if for every $\sigma : [n] \to [r]$, there exists a completion C of Q_σ, such that C reduces to A.*

Theorem 4.16 *The following two implications are valid:*

- *A pair $Q \hookrightarrow A$, where $Q = GLS[n, T]$, $A \in [r]^{l \times l}$, yields a two server r-ary PIR protocol with communication* $\log T$ *from $\mathcal{U}$ to each $\mathcal{S}_j$ and communication* $\log l$ *from $\mathcal{S}_j$'s back to $\mathcal{U}$.*

- *A two server r-ary PIR protocol with queries of length $t(n)$ and answers of length $a(n)$, where the user tosses at most $\tau(n)$ random coins yields a pair $Q \hookrightarrow A$, where $Q = GLS\left[n, nr^{t(n)+\tau(n)}\right]$, and A is a r-ary square matrix of size $nr^{t(n)+a(n)}$.*

Proof: We start with the first part. We assume that matrix A is known to all parties $\mathcal{U}, \mathcal{S}_1$ and $\mathcal{S}_2$. At the preprocessing stage, servers use the database $x \in [r]^n$ to define the map $\sigma : [n] \to [r]$, setting $\sigma(i) \stackrel{\mathrm{def}}{=} x_i$. Also, they find an appropriate completion C, and fix maps $\pi_1 : [T] \to [l]$ and $\pi_2 : [T] \to [l]$, such for all j,k : $C_{jk} = A_{\pi_1(j),\pi_2(k)}$. Next, the following protocol is executed.

$\mathcal{U}$	:	Picks a location j,k in Q such that $Q_{jk} = i$ uniformly at random.
$\mathcal{U} \to \mathcal{S}_1$	:	j
$\mathcal{U} \to \mathcal{S}_2$	:	k
$\mathcal{U} \leftarrow \mathcal{S}_1$	:	$\pi_1(j)$
$\mathcal{U} \leftarrow \mathcal{S}_2$	:	$\pi_2(k)$
$\mathcal{U}$	:	Outputs $A_{\pi_1(j),\pi_2(k)}$.

It is straightforward to verify that the protocol above is private, since a uniformly random choice of a location j,k such that $Q_{jk} = i$, induces uniformly random individual distributions on j and on k. Correctness follows from the fact that C reduces to A. Total communication is given by $2(\log T + \log l)$.

Now we proceed to the second part. Consider a two server protocol $\mathcal{P} = (\mathcal{Q}, \mathcal{A}, \mathcal{C})$. First we show that one can modify $\mathcal{P}$ to obtain a new PIR protocol $\mathcal{P}' = (\mathcal{Q}', \mathcal{A}', \mathcal{C}')$, such that $\mathcal{C}'$ depends only on ans'_1 and ans'_2, but not on i or *rand*. The transformation is simple:

- First $\mathcal{Q}'$ obtains a random string *rand* and invokes $\mathcal{Q}(i, rand)$ to generate (que_1, que_2). Next $\mathcal{Q}'$ tosses $\log n$ extra random coins to represent i as a random sum $i = i_1 + i_2 \mod (n)$, sets $que'_1 = que_1 \circ i_1$, $que'_2 = que_2 \circ i_2$ and sends que'_1 to $\mathcal{S}_1$ and que'_2 to $\mathcal{S}_2$.

- For $j = 1,2$ $\mathcal{A}'$ extracts que_j from que'_j, runs $\mathcal{A}$ on (j, x, que_j) and returns $ans_j \circ que'_j$.

- Finally, $\mathcal{C}'$ extracts $que_1, que_2, ans_1, ans_2$ and i from ans'_1 and ans'_2 and performs a brute force search over all possible random coin tosses of $\mathcal{Q}$ to find some random input $rand'$ such that $\mathcal{Q}(i, rand') = (que_1, que_2)$. $\mathcal{C}'$ runs $\mathcal{C}$ on $(ans_1, ans_2, i, rand')$ and returns the answer. Note that the string $rand'$ may in fact be different from the string $rand$ however the correctness property of $\mathcal{P}$ implies that even in this case $\mathcal{C}'$ outputs the right value.

Now consider the protocol $\mathcal{P}'$. Let Q'_j denote the range of queries to server j, and A'_j denote the range of answers from server j. Variable que'_j ranges over Q'_j, and variable ans'_j ranges over A'_j. Let $R(que'_j, i)$ denote the set of random strings $rand$ that lead to query que'_j to server j on input i. Formally,

$$R(que'_1, i) = \{rand \in [r]^{\tau(n)} \mid \exists\ que'_2 : Q(i, rand) = (que'_1, que'_2)\}$$
$$R(que'_2, i) = \{rand \in [r]^{\tau(n)} \mid \exists\ que'_1 : Q(i, rand) = (que'_1, que'_2)\}$$

Note that the privacy property of the protocol $\mathcal{P}'$ implies that the cardinalities of $R(que'_j, i)$ are independent of i. We denote these cardinalities by $rand(que'_j)$. It is easy to see that $rand(que'_j)$ is always an integer between 1 and $r^{\tau(n)}$. Now we are ready to define matrices Q and A.

Rows of Q are labelled by pairs (que'_1, s_1), where $s_1 \in [rand(que'_1)]$. Columns of Q are labelled by pairs (que'_2, s_2), where $s_2 \in [rand(que'_2)]$. We set $Q_{(que'_1,s_1),(que'_2,s_2)} = i$ if there exists a string $rand \in R(que'_1, i) \cap R(que'_2, i)$ such that $rand$ is the string number s_1 in $R(que'_1, i)$ and the string number s_2 in $R(que'_2, i)$ with respect to lexicographic ordering of these sets; otherwise we set $Q_{(que'_1,s_1),(que'_2,s_2)} = *$.

Consider an arbitrary pair $(i, (que'_1, s_1))$, where $s_1 \in [rand(que'_1)]$. Let $rand$ be the random string number s_1 in lexicographic ordering of $R(que'_1, i)$. Let $\mathcal{Q}'(i, rand) = (que'_1, que'_2)$, and let s_2 be the number of $rand$ in lexicographic ordering of $R(que'_2, i)$. The column of Q labelled (que'_2, s_2) is the unique column such that $Q_{(que'_1,s_1),(que'_2,s_2)} = i$. Thus we proved that every row of Q contains exactly one entry labelled i. A similar argument proves this claim for columns. Thus Q is a generalized Latin square.

Now we proceed to matrix A. Rows of A are labelled by possible values of ans'_1, similarly columns of A are labelled by possible values of ans'_2. We set $A_{ans'_1,ans'_2} = \mathcal{C}'(ans'_1, ans'_2)$. The unspecified entries of A are set arbitrarily. Matrix A defined above may not be a square, however one can always pad it to a square shape.

It remains to show that $Q \hookrightarrow A$. Given a map $\sigma : [n] \to [r]$ we consider a database

x, where $x_i = \sigma(i)$. We use protocol $\mathcal{P}'$ to define maps π_1 from the row set of Q to the row set of A, and π_2 from the column set of Q to the column set of A. We set $\pi_1(que'_1, s_1) = \mathcal{A}'(1, x, que'_1)$ and $\pi_2(que'_2, s_2) = \mathcal{A}'(2, x, que'_2)$. Correctness property of $\mathcal{P}'$ implies that maps π_1, π_2 reduce certain completion of Q_σ to A. ■

The theorem above represents our combinatorial view of two server PIR protocols. A PIR protocol is just a pair $Q \hookrightarrow A$, where Q is a generalized Latin square and A is a r-ary matrix. Every PIR protocol can be converted into this form, and in case the number of user's coin tosses is linear in the query length such conversion does not affect the asymptotic communication complexity.

4.3.1 Bilinear PIR

The combinatorial interpretation of PIR suggested above views PIR as a problem of reducing certain special families of matrices to some fixed matrix. A nice example of a nontrivial matrix where one can say a lot about matrices that reduce to it is a Hadamard matrix. In what follows we assume that the alphabet size r is a prime power.

Definition 4.17 *A Hadamard matrix H_t is a r^t by r^t matrix whose rows and columns are labelled by elements of $\mathbb{F}_r^t$ and matrix cells contain the dot products of the corresponding labels, i.e., $(H_t)_{v_1,v_2} = (v_1, v_2)$.*

Lemma 4.18 *Let M be a square matrix with entries from $\mathbb{F}_r$; then M reduces to the Hadamard matrix H_t if and only if the rank of M is at most t.*

Proof: Clearly, the rank of H_t is t therefore the rank of any matrix that reduces to H_t is at most that large. To prove the converse observe that M can be written as a sum of t matrices $M = M^1 + \ldots + M^t$, where each M^j is of rank at most one. Let m be the size of M. For every $i \in [t]$ set the i-th coordinate of t long vectors $v^1, \ldots, v^m$ $u^1, \ldots, u^m$ so that $v_i^j u_i^k = M_{jk}^i$. Now the maps $\pi_1 : [m] \to [r^t]$, $\pi_2 : [m] \to [r^t]$ defined by $\pi_1(j) = v^j$, $\pi_2(k) = u^k$ embed M into H_t. ■

The above lemma is important since it allows to reduce the proof that $Q \hookrightarrow H_t$ for some generalized Latin square Q to showing that for every $\sigma : [n] \to \mathbb{F}_r$, Q_σ can be completed to a low rank matrix.

Definition 4.19 *We say that a two server PIR scheme $Q \hookrightarrow A$ is bilinear if $A = H_t$ for some value of t.*

Another way to formulate the above definition is to say that a PIR scheme is bilinear if $\mathcal{U}$ computes the dot product of servers' answers to obtain value of x_i.

4.3.2 Group based PIR

Finite groups are a natural source of generalized Latin squares $Q = \mathrm{GLS}[n, T]$. Let $G = \{g_1, \dots, g_T\}$ be a finite group of size T. Let $S = \{s_1, \dots, s_n\} \subseteq G$ be an ordered subset of G of size n. A generalized Latin square $Q_{G,S}$ is a T by T square matrix whose rows and columns are labelled by elements of G, and $Q_{g_1,g_2} = i$ if $g_1 g_2^{-1} = s_i$, while all other locations contain stars.

When PIR protocol $Q \hookrightarrow A$ uses a generalized Latin square $Q_{G,S}$ we say that such protocol *employs a group based secret sharing scheme.* Essentially, this means that given an index i, $\mathcal{U}$ maps it to a group element s_i, represents s_i as a random product in the group $s_i = g_1 g_2^{-1}$ and sends g_j to $\mathcal{S}_j$.

The notion of a *group based* PIR protocol (for that we later prove a lower bound) is more restrictive. Let $M \in [r]^{T \times T}$ and G be finite group. Assume that rows and columns of M are labelled by $g_1, \dots, g_T$. We say that M *respects* G if for every $g_1, g_2, g_3, g_4 \in G$ such that $g_1 g_2^{-1} = g_3 g_4^{-1}$, we have $M_{g_1,g_2} = M_{g_3,g_4}$.

Definition 4.20 *We say that PIR protocol $Q \hookrightarrow A$ is group based if it employs a secret sharing scheme based on some group G and for every $\sigma : [n] \to \mathbb{F}_r$ there exists a completion C such that C reduces to A and C respects G.*

Stated in other words a PIR scheme is group based if servers represent database by a function on a certain finite group G and the scheme allows user to retrieve the value of this function at any group element using the natural secret sharing based on G.

4.4 Complexity of bilinear group based PIR

Consider a bilinear group based PIR scheme $Q \hookrightarrow H_t$ based on a group G, with answer length t. Clearly, query length is $\log |G|$. Let $A(r, G, t)$ denote the number of $|G|$ by $|G|$ matrices over $\mathbb{F}_r$ that respect G (for some fixed labelling $\{g_1, \ldots, g_T\}$ or rows and columns) and have rank at most t. It is easy to see that

$$r^n \leq A(r, G, t), \tag{4.1}$$

since by lemma 4.18 every database yields such a matrix and distinct databases yield distinct matrices. In subsection 4.4.2 we obtain an equivalent algebraic definition for $A(r, G, t)$, and in subsection 4.4.3 we prove an upper bound for $A(r, G, t)$. Our final result is a constraint on the range of possible values of $r, |G|, t$. This constraint implies an $\Omega(n^{1/3})$ lower bound for total communication of any bilinear group based PIR scheme.

4.4.1 Algebraic preliminaries

Our proof relies on some basic notions of representation theory of finite groups. The standard references for this subject are [Wei03], [Isa76].

Let $G = \{g_1, \ldots, g_T\}$ be a finite (not necessarily abelian) group. General linear group $GL_t(\mathbb{F}_r)$ is a multiplicative group of all non-degenerate t by t matrices over $\mathbb{F}_r$.

- An $\mathbb{F}_r$-*representation* of G of degree t is an homomorphism $\phi : G \rightarrow GL_t(\mathbb{F}_r)$.
- A group algebra $\mathbb{F}_r[G]$ of G over a field $\mathbb{F}_r$ is an algebra over $\mathbb{F}_r$ consisting of all possible formal linear combinations $\sum_{i=1}^{T} \alpha_i g_i$, $\alpha_i \in \mathbb{F}_r$. The algebraic operations in $\mathbb{F}_r[G]$ are defined by:

$$\begin{aligned}\sum_i \alpha_i g_i + \sum_i \beta_i g_i &= \sum_i (\alpha_i + \beta_i) g_i;\\ \left(\sum_i \alpha_i g_i\right) * \left(\sum_i \beta_i g_i\right) &= \sum_{i,j} (\alpha_i \beta_j)(g_i g_j);\\ \lambda \left(\sum_i \alpha_i g_i\right) &= \sum_i (\lambda \alpha_i) g_i, \quad \lambda \in \mathbb{F}_r.\end{aligned}$$

- For an algebra A over $\mathbb{F}_r[G]$, a left A-module is an $\mathbb{F}_r$-linear space on which A acts by left multiplication in such a way that for any $m_1, m_2 \in M$ and any $\alpha, \beta \in \mathbb{F}_r[G]$:

  $$\alpha(m_1 + m_2) = \alpha m_1 + \alpha m_2;$$
  $$(\alpha + \beta)m_1 = \alpha m_1 + \beta m_1;$$
  $$(\alpha\beta)m_1 = \alpha(\beta m_1).$$

  Dimension of a module is its dimension as an $\mathbb{F}_r$-linear space. Two A-modules are called isomorphic if there exists an isomorphism between them as linear spaces that preserves multiplication by the elements of A.

- There is a one to one correspondence between t dimensional left $\mathbb{F}_r[G]$-modules M considered up to isomorphism and $\mathbb{F}_r$-representations of G of degree t considered up to inner automorphisms of the $GL_t(\mathbb{F}_r)$.

4.4.2 Algebraic formulation

Let $A = \mathbb{F}_r[G]$. For $\alpha \in A$, let $\text{rk}(\alpha) = \dim(A\alpha)$, where $\dim(A\alpha)$ is the dimension of $A\alpha$ as a linear space over $\mathbb{F}_r$. Consider the *regular representation* ϕ of G, $\phi : G \longrightarrow GL_{|G|}(\mathbb{F}_r)$, defined by

$$(\phi(g))_{g_1,g_2} = \begin{cases} 1, & g_1 g_2^{-1} = g, \\ 0, & \text{otherwise.} \end{cases} \tag{4.2}$$

Extend ϕ to A by linearity. Note that ϕ is an injective algebra homomorphism and that image of ϕ is the $\mathbb{F}_r$-algebra R of all matrices that respect G. Observe that for any $M \in R$,

$$\text{rk}M = \dim\{M'M \mid M' \in R\}. \tag{4.3}$$

To verify formula (4.3) one needs to notice that the first row of a matrix $M' \in R$ can be arbitrary. Therefore products $M'M$ contain all possible linear combinations of rows of M as their first row. Also notice that matrices in R are uniquely determined by their first row. Formula (4.3) follows. It implies an algebraic definition for $A(r, G, t)$:

$$A(r, G, t) = \#\{\alpha \in \mathbb{F}_r[G] \mid \text{rk}(\alpha) \leq t\}. \tag{4.4}$$

4.4.3 Low dimensional principal ideals in group algebras

Let V be an $\mathbb{F}_r$-linear subspace of A. Left annihilator of V is defined by $Ann_L(V) \stackrel{\text{def}}{=} \{\beta \in A \mid \beta V = 0\}$. Similarly, right annihilator $Ann_R(V) \stackrel{\text{def}}{=} \{\beta \in A \mid V\beta = 0\}$. Clearly, $Ann_L(V)$ is a left ideal in A and $Ann_R(V)$ is a right ideal in A. Let M be a left A-module. Kernel of M is defined by $Ker(M) \stackrel{\text{def}}{=} \{\beta \in A \mid \beta M = 0\}$. It is straightforward to verify that $Ker(M)$ is a two sided ideal that coincides with $Ann_L(M)$ if M is a left ideal in A.

Lemma 4.21 *The number of t-dimensional left A-modules counted up to isomorphism is at most $r^{\log|G|t^2}$.*

Proof: The fourth bullet from subsection 4.4.1 implies that it suffices to count $\mathbb{F}_r$-representations of G of degree t. Let $g_1, \dots, g_s$ be the set of generators for G, where $s \leq \log|G|$. Now we only have to note that every representation $\phi : G \to GL_t(\mathbb{F}_r)$ is uniquely specified by s matrices $\phi(g_1), \dots, \phi(g_s)$ each of size t by t. ■

Clearly, isomorphic modules have identical kernels. Now we show that kernel of a low dimensional module has high dimension.

Lemma 4.22 *Let M be an t-dimensional left A-module; then the dimension of $Ker(M)$ as an $\mathbb{F}_r$-linear space is at least $|G| - t^2$.*

Proof: Note that multiplication by an element of A induces a linear transformation of M. Such transformation can be expressed by an t by t matrix. Multiplication by a linear combination of elements of A corresponds to linear combination of corresponding matrices. Therefore we conclude that $\dim Ker(M) \geq |G| - t^2$. ■

Lemma 4.23 *Suppose V is an $\mathbb{F}_r$-linear subspace of A; then* $\dim(Ann_R(V)) \leq |G| - \dim(V)$.

Proof: Consider a bilinear map $l : A \otimes A \to \mathbb{F}_r$, setting $l(x \otimes y)$ equal to the coefficient of 1 in the expansion of xy in the group basis. Clearly, l has full rank (since in the group basis l is defined by an identity matrix up to a permutation of columns). However $l(V \otimes Ann_R(V)) = 0$. Thus $\dim(Ann_R(V)) \leq |G| - \dim(V)$. ■

Our main technical result is given by

Theorem 4.24 *For arbitrary finite group G and arbitrary values of r and t*

$$A(r, G, t) \leq r^{O(\log |G| t^2)}.$$

Proof: Let $\alpha \in A$ be such that $\mathrm{rk}(\alpha) \leq t$. Consider $A\alpha$ as a left A-module. $Ker(A\alpha)$ is a two-sided ideal $I = Ann_L(A\alpha)$. Note that $\alpha \in Ann_R(I)$. By lemma 4.21 every A-module of dimension up to t has its kernel coming from a family of at most $tr^{\log |G| t^2}$ ideals. Also by lemmas 4.22 and 4.23 there are at most r^{t^2} elements in $Ann_R(I)$ for every I. ■

Combining equation (4.1) with theorem 4.24 we obtain our main result.

Theorem 4.25 *Let $Q \hookrightarrow H_t$ be a bilinear group based PIR scheme over a group G. Let $q = \log |G|$ denote the query length and t denote the answer length; then*

$$n \leq O(qt^2).$$

In particular total communication of any such scheme is $\Omega(n^{1/3})$.

4.5 Summary of lower bounds for two server PIR

In sections 4.3–4.5 we introduced a novel, though quite natural combinatorial view of the two server PIR problem, and obtained a lower bound for communication complexity of PIR in a restricted (bilinear group-based) model, that captures all currently known PIR protocols [RY06, appendix]. Stated informally, our main result is that as long as servers represent database by a function on a finite group, protocol allows user to retrieve the value of this function at any group element, and user computes the dot product of servers responses to obtain the final answer communication complexity has to be $\Omega(n^{1/3})$.

Clearly, our lower bound admits two interpretations. On the one hand it can be viewed as a witness in support of conjecture of Chor et al. from [CGKS98] saying that their PIR protocol with $O(n^{1/3})$ communication is asymptotically optimal. On the other hand our result exhibits a common shortcoming of the existing upper bound techniques and thus hopefully may provide some directions for future work on upper bounds.

Bibliography

[AFK89] Martin Abadi, Joan Feigenbaum, and Joe Kilian. On hiding information from an oracle. *Journal of Computer and System Sciences*, 39:21–50, 1989.

[AKS04] Manindra Agrawal, Neeraj Kayal, and Nitin Saxena. PRIMES is in P. *Annals of Mathematics*, 160:781–793, 2004.

[Amb97] Andris Ambainis. Upper bound on the communication complexity of private information retrieval. In *32th International Colloquium on Automata, Languages and Programming (ICALP)*, volume 1256 of Lecture Notes in Computer Science, pages 401–407, 1997.

[BB65] Edwin Beckenbach and Richard Bellman. *Inequalities.* 1965.

[BC06] Jean Bourgain and Mei-Chu Chang. A Gauss sum estimate in arbitrary finite fields. *Comptes Rendus Mathematique*, 342:643–646, 2006.

[BF90] Donald Beaver and Joan Feigenbaum. Hiding instances in multioracle queries. In *7th International Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 415 of Lecture Notes in Computer Science, pages 37–48, 1990.

[BFG06] Richard Beigel, Lance Fortnow, and William Gasarch. A tight lower bound for restricted PIR protocols. *Computational Complexity*, 15:82–91, 2006.

[BFKR90] Donald Beaver, Joan Feigenbaum, Joe Kilian, and Phillip Rogaway. Security with low communication overhead. In *International Cryptology Conference (CRYPTO)*, pages 62–76, 1990.

[BFLS91] Laszlo Babai, Lance Fortnow, Leonid Levin, and Mario Szegedy. Checking computations in polylogarithmic time. In *23th ACM Symposium on Theory of Computing (STOC)*, pages 21–31, 1991.

[BFNW93] Laszlo Babai, Lance Fortnow, Naom Nisan, and Avi Wigderson. BPP has subexponential time simulations unless EXPTIME has publishable proofs. *Computational Complexity*, 3:307–318, 1993.

[BIK05] Amos Beimel, Yuval Ishai, and Eyal Kushilevitz. General constructions for information-theoretic private information retrieval. *Journal of Computer and System Sciences*, 71:213–247, 2005.

[BIKM99] Amos Beimel, Yuval Ishai, Eyal Kushilevitz, and Tal Malkin. One-way functions are essential for single-server private information retrieval. In *31nd ACM Symposium on Theory of Computing (STOC)*, pages 89–98, 1999.

[BIKR02] Amos Beimel, Yuval Ishai, Eyal Kushilevitz, and Jean-Francios Raymond. Breaking the $O\left(n^{1/(2k-1)}\right)$ barrier for information-theoretic private information retrieval. In *43rd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 261–270, 2002.

[BIM00] Amos Beimel, Yuval Ishai, and Tal Malkin. Reducing the servers' computation in private information retrieval: PIR with preprocessing. In *International Cryptology Conference (CRYPTO)*, volume 1880 of Lecture Notes in Computer Science, pages 56–74, 2000.

[BIW07] Omer Barkol, Yuval Ishai, and Enav Weinreb. On locally decodable codes, self-correctable codes, and t-private PIR. 2007.

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47:549–595, 1993.

[BOGW88] Michael Ben-Or, Shafi Goldwasser, and Avi Wigderson. Completeness theorems for non-cryptographic fault-tolerant distributed computation. In *20th ACM Symposium on Theory of Computing (STOC)*, pages 1–10, 1988.

[BS02] Amos Beimel and Yoav Stahl. Robust information theoretic private information retrieval. In *3rd Conference of Security in Communication Networks*, 2002.

[CB] Curtis Cooper and Steven Boone. http://www.mersenne.org/32582657.htm.

[CCD88] David Chaum, Claude Crepeau, and Ivan Damgard. Multiparty unconditionally secure protocols. In *20th ACM Symposium on Theory of Computing (STOC)*, pages 11–19, 1988.

[CGKS98] Benny Chor, Oded Goldreich, Eyal Kushilevitz, and Madhu Sudan. Private information retrieval. *Journal of the ACM*, 45:965–981, 1998.

[CIK$^+$01] Ran Canetti, Yuval Ishai, Ravi Kumar, Michael Reiter, Ronitt Rubinfeld, and Rebecca Wright. Selective private function evaluation with applications to private statistics. In *20th ACM Symposium on Principles of Distributed Computing (PODC)*, pages 293–304, 2001.

[CLO96] David Cox, John Little, and Donal O'Shea. *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra.* 1996.

[CMS99] Christian Cachin, Silvio Micali, and Markus Stadler. Computationally private information retrieval with polylogarithmic communication. In *International Cryptology Conference (EUROCRYPT)*, volume 1592 of Lecture Notes in Computer Science, pages 402–414, 1999.

[DBR90] Silvio Micali Donald Beaver and Pillip Rogaway. The round complexity of secure protocols. In *22nd ACM Symposium on Theory of Computing (STOC)*, pages 503–513, 1990.

[DCIO01] Giovanni Di-Crescenzo, Yuval Ishai, and Rafail Ostrovsky. Universal service-providers for private information retrieval. *Journal of Cryptology*, 14:37–74, 2001.

[DCMO00] Giovanni Di-Crescenzo, Tal Malkin, and Rafail Ostrovsky. Single-database private information retrieval implies oblivious transfer. In *International Cryptology Conference (EUROCRYPT)*, volume 1807 of Lecture Notes in Computer Science, pages 122–138, 2000.

[DJK$^+$02] A. Deshpande, R. Jain, T. Kavitha, S. Lokam, and J. Radhakrishnan. Better lower bounds for locally decodable codes. In *20th IEEE Computational Complexity Conference (CCC)*, pages 184–193, 2002.

[dW03] B.L. Van der Waerden. *Algebra.* 2003.

[ES76] P. Erdos and T. Shorey. On the greatest prime factor of $2^p - 1$ for a prime p and other expressions. *Acta. Arith.*, 30:257–265, 1976.

[FLN$^+$02] Eldar Fischer, Eric Lehman, Ilan Newman, Sofya Raskhodinkova, Ronitt Rubinfeld, and Alex Samorodnitsky. Monotonicity testing over the general poset domains. In *34th ACM Symposium on Theory of Computing (STOC)*, pages 474–483, 2002.

[FLS07] Kevin Ford, Florian Luca, and Igor Shparlinski. On the largest prime factor of the Mersenne numbers. Arxiv 0704.1327, April 2007.

[Gas04] William Gasarch. A survey on private information retrieval. *The Bulletin of the EATCS*, 82:72–107, 2004.

[GGM98] Yael Gertner, Shafi Goldwasser, and Tal Malkin. A random server model for private information retrieval. In *International Workshop on Random-*

ization and Computation (RANDOM), volume 1518 of Lecture Notes in Computer Science, pages 200–217, 1998.

[GIKM00] Yael Gertner, Yuval Ishai, Eyal Kushilevitz, and Tal Malkin. Protecting data privacy in private information retrieval schemes. *Journal of Computer and System Sciences*, 60:592–629, 2000.

[GKST02] Oded Goldreich, Howard Karloff, Leonard Schulman, and Luca Trevisan. Lower bounds for locally decodable codes and private information retrieval. In *17th IEEE Computational Complexity Conference (CCC)*, pages 175–183, 2002.

[Gol05] Oded Goldreich. Short locally testable codes and proofs. Electronic Colloquium on Computational Complexity (ECCC) TR05-014, 2005.

[GR05] Craig Gentry and Zulfikar Ramzan. Single-database private information retrieval with constant communication rate. In *32th International Colloquium on Automata, Languages and Programming (ICALP)*, pages 803–815, 2005.

[IK04] Yuval Ishai and Eyal Kushilevitz. On the hardness of information-theoretic multiparty computation. In *Eurocrypt 2004*, volume 3027 of Lecture Notes in Computer Science, pages 439–455, 2004.

[Isa76] I. Martin Isaacs. *Character theory of finite groups.* 1976.

[Ito99] Toshiya Itoh. Efficient private information retrieval. *IEICE Trans. Fund. of Electronics, Commun. and Comp. Sci.*, pages 11–20, 1999.

[Ito01] Toshiya Itoh. On lower bounds for the communication complexity of private information retrieval. *IEICE Trans. Fund. of Electronics, Commun. and Comp. Sci.*, pages 157–164, 2001.

[KdW04] Iordanis Kerenidis and Ronald de Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. *Journal of Computer and System Sciences*, 69:395–420, 2004.

[KO97] Eyal Kushilevitz and Rafail Ostrovsky. Replication is not needed: Single-database computationally-private information retrieval. In *38rd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 364–373, 1997.

[KO00] Eyal Kushilevitz and Rafail Ostrovsky. One-way trapdoor permutations are sufficient for single-database computationally-private information retrieval. In *International Cryptology Conference (EUROCRYPT)*, volume 1807 of Lecture Notes in Computer Science, pages 104–121, 2000.

[KT00] Jonathan Katz and Luca Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In *32th ACM Symposium on Theory of Computing (STOC)*, pages 80–86, 2000.

[KY01] Aggelos Kiayias and Moti Yung. Secure games with polynomial expressions. In *28th International Colloquium on Automata, Languages and Programming (ICALP)*, volume 2076 of Lecture Notes in Computer Science, pages 939–950, 2001.

[KY07] Kiran S. Kedlaya and Sergey Yekhanin. Locally decodable codes from nice subsets of finite fields and prime factors of Mersenne numbers. Electronic Colloquium on Computational Complexity (ECCC) TR07-040, 2007.

[Lip90] Richard Lipton. Efficient checking of computations. In *7th International Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 415 of Lecture Notes in Computer Science, pages 207–215, 1990.

[Lip04] Helger Lipmaa. An oblivious transfer protocol with log-squared communication. International Association for Cryptologic Research Technical Report 2004/063, 2004.

[LN83] Rudolf Lidl and Harald Niederreiter. *Finite Fields.* 1983.

[LPW] Lenstra-Pomerance-Wagstaff conjecture. In Wikipedia, The Free Encyclopedia.

[Man98] Eran Mann. Private access to distributed information. Master's thesis, Technion - Israel Institute of Technology, Haifa, 1998.

[Mer44] Marin Mersenne. *Cogitata Physica-Mathematica.* 1644.

[MP04] Leo Murata and Carl Pomerance. On the largest prime factor of a Mersenne number. *Number theory, CRM Proc. Lecture Notes of American Mathematical Society*, 36:209–218, 2004.

[MS77] F.J. MacWilliams and N.J.A. Sloane. *The theory of error correcting codes.* 1977.

[MW00] M. Murty and S. Wong. The ABC conjecture and prime divisors of the Lucas and Lehmer sequences. In *Milennial Conference on Number Theory III*, pages 43–54, Urbana, IL, 2000.

[NP99] Moni Naor and Benny Pinkas. Oblivious transfer and polynomial evaluation. In *29th ACM Symposium on Theory of Computing (STOC)*, pages 245–254, 1999.

[Oba02] Kenji Obata. Optimal lower bounds for 2-query locally decodable linear codes. In *6th International Workshop on Randomization and Computation (RANDOM)*, volume 2483 of Lecture Notes in Computer Science, pages 39–50, 2002.

[OS97] Rafail Ostrovsky and Victor Shoup. Private information storage. In *29th ACM Symposium on Theory of Computing (STOC)*, pages 294–303, 1997.

[OS07] Rafail Ostrovsky and William Skeith. A survey of single database PIR: techniques and applications. Cryptology ePrint Archive Report 059, 2007.

[Pom80] Carl Pomerance. Recent developments in primality testing. *Math. Intelligencer*, 3:97–105, 1980.

[Pri] The prime pages. http://primes.utm.edu/.

[PS94] Alexander Polishchuk and Daniel Spielman. Nearly-linear size holographic proofs. In *26th ACM Symposium on Theory of Computing (STOC)*, pages 194–203, 1994.

[Rag07] Prasad Raghavendra. A note on Yekhanin's locally decodable codes. Electronic Colloquium on Computational Complexity (ECCC) TR07-016, 2007.

[Rom06] Andrei Romashchenko. Reliable computations based on locally decodable codes. In *23rd International Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 3884 of Lecture Notes in Computer Science, pages 537–548, 2006.

[Ros88] M. Rosen. A proof of the Lucas-Lehmer test. *American Mathematical Monthly*, 95:855–856, 1988.

[RS78] Imre Ruzsa and Endre Szemeredi. Triple systems with no six points carrying three trinangels. *Colloquia Mathematica Societatis Janos Bolyai*, 18:939–945, 1978.

[RY06] Alexander Razborov and Sergey Yekhanin. An $\Omega(n^{1/3})$ lower bound for bilinear group based private information retrieval. In *47rd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 739–748, 2006.

[Sar] Peter Sarnak. Personal commuication.

[Sch62] A. Schinzel. On primitive factors of $a^n - b^n$. In *Cambridge Philos. Soc.*, volume 58, pages 555–562, 1962.

[SS05] Gabor Sarkozy and Stanley Selkow. An extension to the Ruzsa-Szemeredi theorem. *Combinatorica*, 25:77–84, 2005.

[Ste74] C. Stewart. The greatest prime factor of $a^n - b^n$. *Acta Arith.*, 26:427–433, 1974.

[Ste77] C. Stewart. On divisors of Fermat, Fibonacci, Lucas, and Lehmer numbers. In *London Math. Soc.*, volume 35, pages 425–447, 1977.

[Ste98] Julien Stern. A new and efficient all-or-nothing disclosure of secrets protocol. In *International Cryptology Conference (ASIACRYPT)*, volume 1514 of Lecture Notes in Computer Science, pages 357–371, 1998.

[STV99] Madhu Sudan, Luca Trevisan, and Salil Vadhan. Pseudorandom generators without the XOR lemma. In *39th ACM Symposium on Theory of Computing (STOC)*, pages 537–546, 1999.

[Sud92] Madhu Sudan. *Efficient checking of polynomials and proofs and the hardness of approximation problems.* PhD thesis, University of California at Berkeley, 1992.

[Tre04] Luca Trevisan. Some applications of coding theory in computational complexity. *Quaderni di Matematica*, 13:347–424, 2004.

[vL82] J.H. van Lint. *Introduction to Coding Theory.* 1982.

[Wag83] Samuel Wagstaff. Divisors of mersenne numbers. *Math. Comp.*, 40:385–397, 1983.

[WdW05] Stephanie Wehner and Ronald de Wolf. Improved lower bounds for locally decodable codes and private information retrieval. In *32nd International Colloquium on Automata, Languages and Programming (ICALP)*, volume 3580 of Lecture Notes in Computer Science, pages 1424–1436, 2005.

[Wei03] S.H. Weintraub. Representation theory of finite groups: algebra and arithmetic. volume 59 of *Graduate studies in mathematics.* AMS, 2003.

[Woo07] David Woodruff. New lower bounds for general locally decodable codes. Electronic Colloquium on Computational Complexity (ECCC) TR07-006, 2007.

[WY05] David Woodruff and Sergey Yekhanin. A geometric approach to information theoretic private information retrieval. In *20th IEEE Computational Complexity Conference (CCC)*, pages 275–284, 2005.

[Yek06] Sergey Yekhanin. New locally decodable codes and private information retrieval schemes. Electronic Colloquium on Computational Complexity (ECCC) TR06-127, 2006.

[Yek07] Sergey Yekhanin. Towards 3-query locally decodable codes of subexponential length. In *39th ACM Symposium on Theory of Computing (STOC)*, pages 266–274, 2007.